CSCI5070 Advanced Topics in Social Computing

Human Computation and Crowdsourcing

Irwin King

The Chinese University of Hong Kong

king@cse.cuhk.edu.hk

CrowdConf 2011

November 1 -2

Mission Bay Conference Center
San Francisco

BUY TICKETS NOW!

SPONSOR
Learn more

VIDEO
CrowdConf2010 Highlights

TICKETS
Register now

CONTACT US

Speakers

Sharon Chiarella
Vice President, Amazon Mechanical Turk

James Dipadua
Senior Program and Product Manager, Servio

Darin Dinsmore
CEO, Crowdbrite

Pia Erkinheimo
Nokia: Ideas Project, Nokia

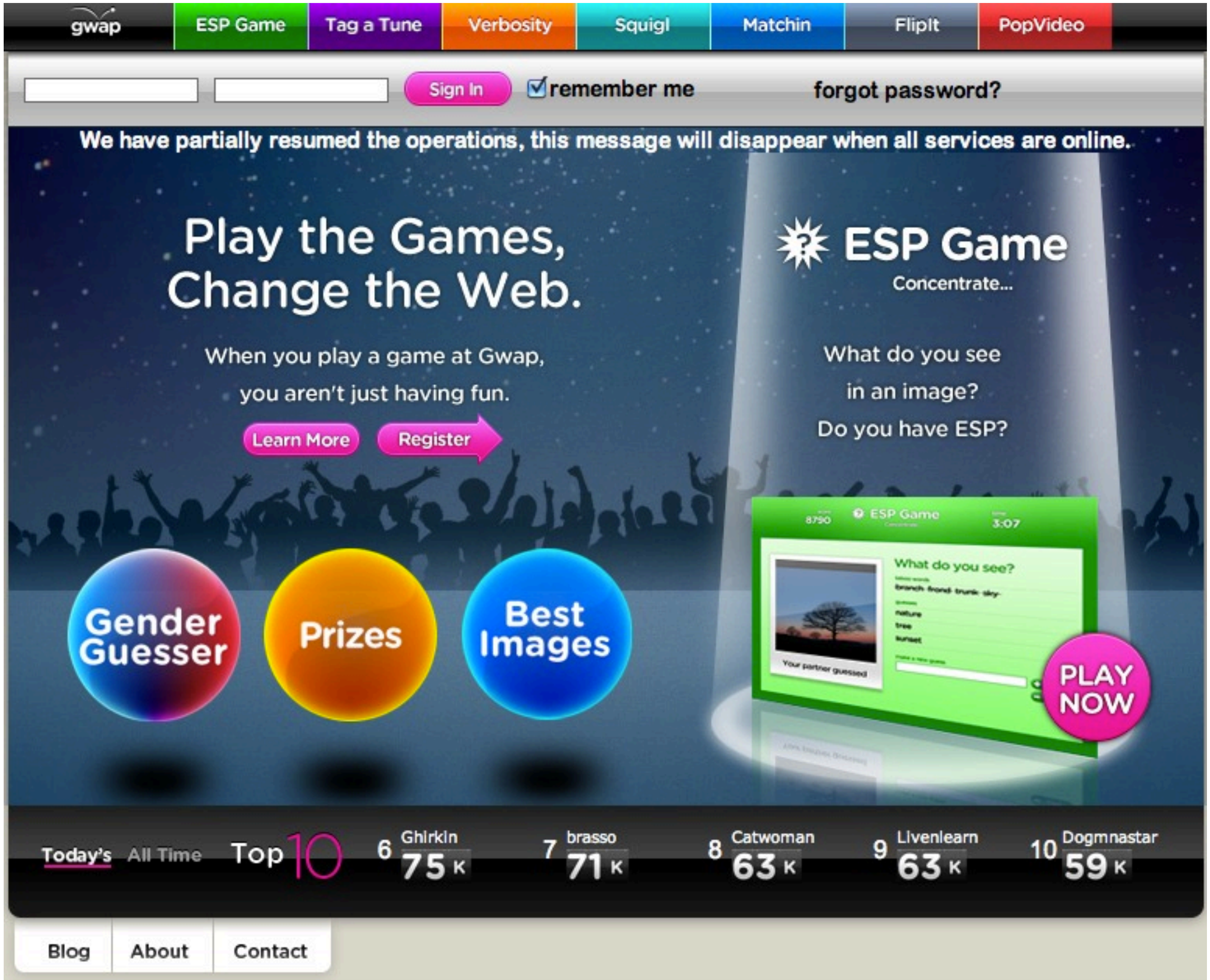
Carl Esposti
Founder, Crowdsourcing.org

Charlie Festa
Community Manager, Threadless

Matt Fisher
Director, Liveops

Michael Gebert
Founder & CEO, Marketing Society

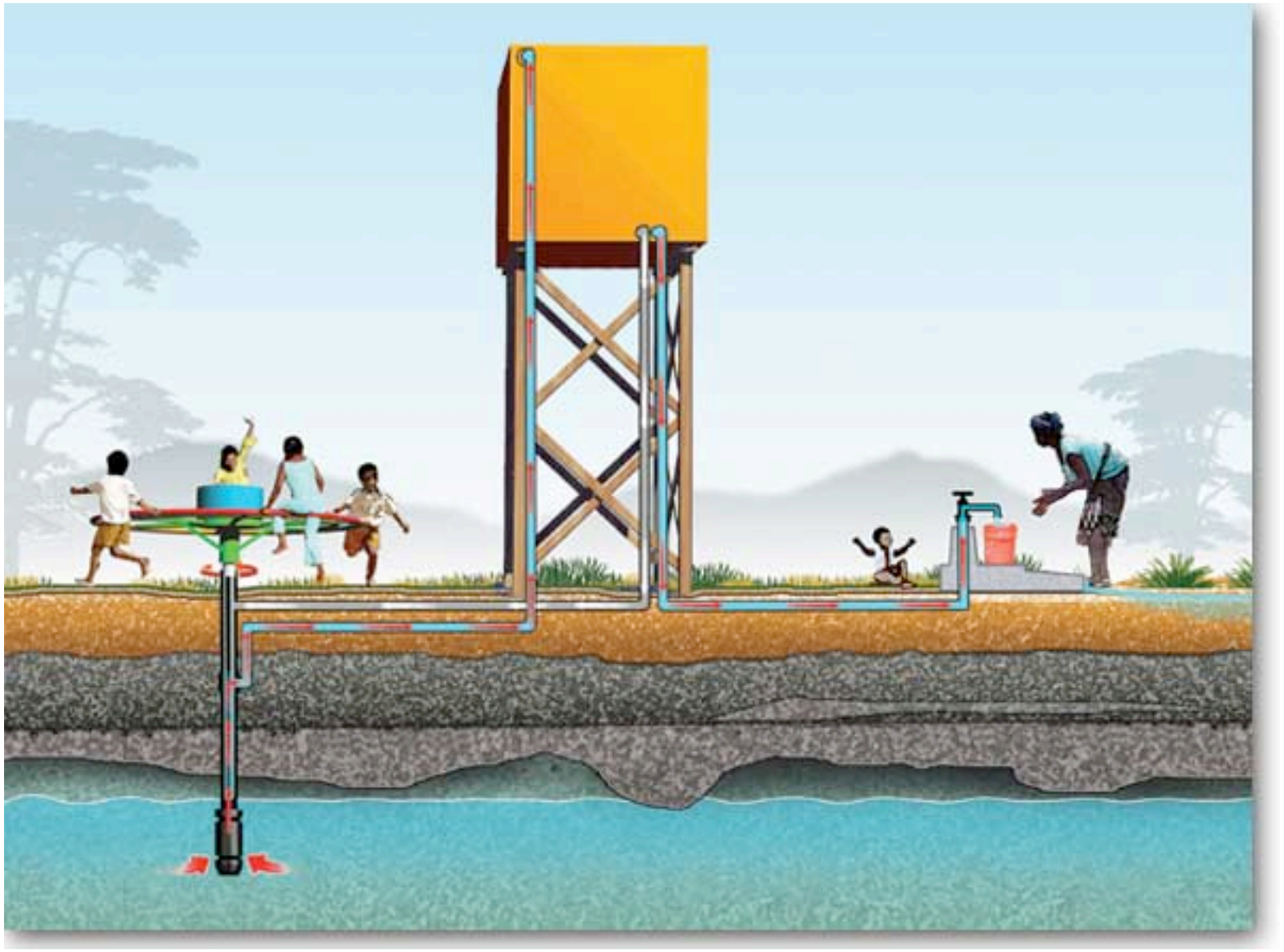
Mark G
Chairman, Lehrman



http://www.gwap.com/

Playing/Having Fun ⟷? Work/Computation

Idea of Human Computation



- Take advantage of people's desire to be entertained and perform useful tasks as a side effect

Motivations

- To describe the categorization of Human Computation Systems (HCS)
- To describe each category of HCS and present the previous work on each category
- To summarize the current state-of-the-art HCS

Why Is It Important?

- Some statistics (July 2008)
  - 200,000+ players have contributed 50+ million labels.
  - Each player plays for a total of 91 minutes.
  - The throughput is about 233 labels/player/hour (i.e., one label every 15 seconds)
- Idea behind
  - Solve some problems which are difficult to be solved by computers.
  - Take advantage of people's desire to be entertained.
  - Produce useful metadata as a by-product.

Outline

- Motivation and Background
- Types of Human Computation
  - Initiatory Human Computation
  - Distributed Human Computation
    - Social Game-based Human Computation with volunteers or paid engineers
    - Social Game-based Human Computation with online players
- Properties of Social Games
- Future Work and Final Remarks

Background

- Human Computation Systems (HCS) aim to solve Artificial Intelligence (AI) problems through the human human interactions

- In order to ensure the collected information to be useful, we have to:

1. guarantee the quality of collected information
2. attract more people to contribute information

Types of HCS

- The categories of the human computation systems are:

1. Initiatory Human Computation
2. Distributed Human Computation
3. Social Game-based Human Computation with volunteers or paid engineers
4. Social Game-based Human Computation with online players

Initiatory Human Computation (1)

- Objective: To complete some tasks that are natural for humans but difficult for computers even computation power increased rapid recently

- Example (1): CAPTCHA
  - A computer generated challenge-response test
  - Objective: To distinguish humans from computers using a common sense problem

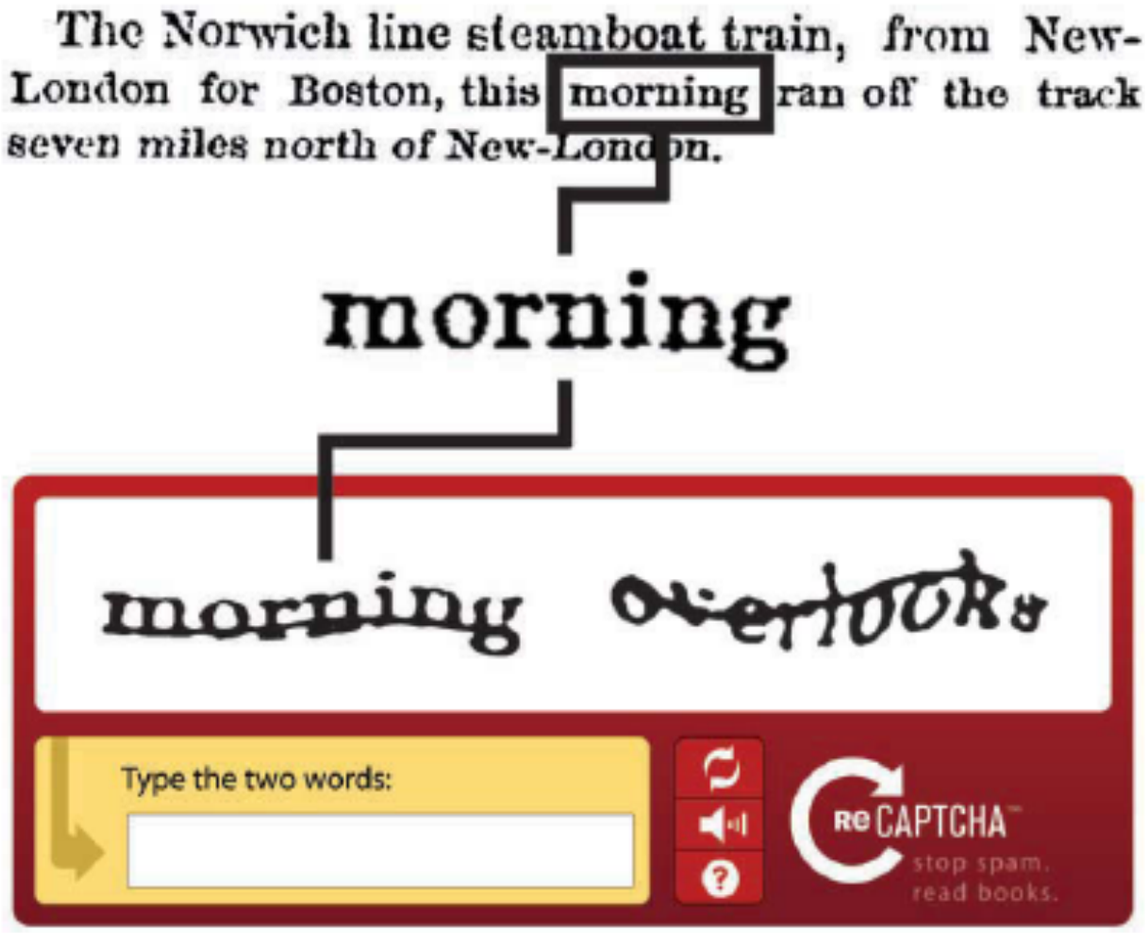
The Yahoo! CAPTCHA.

Initiatory Human Computation (2)

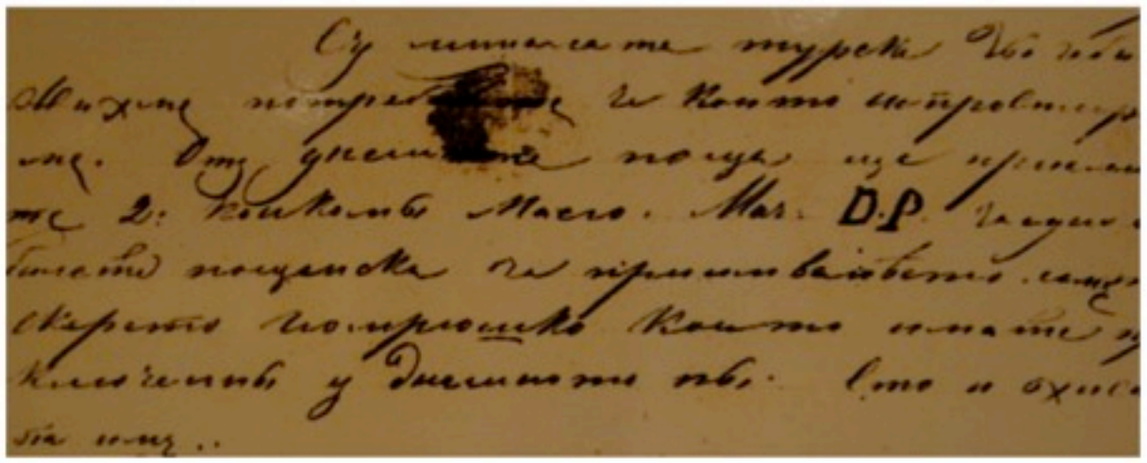
- Example (2): reCAPTCHA
  - Objective: To produce valuable common sense knowledge to improve the OCR quality in digitizing books
  - Combining two words: one identified word; and one unidentified word
  - If a user recognizes the identified word, the answer to the unidentified word is assumed to be correct

Initiatory Human Computation (3)

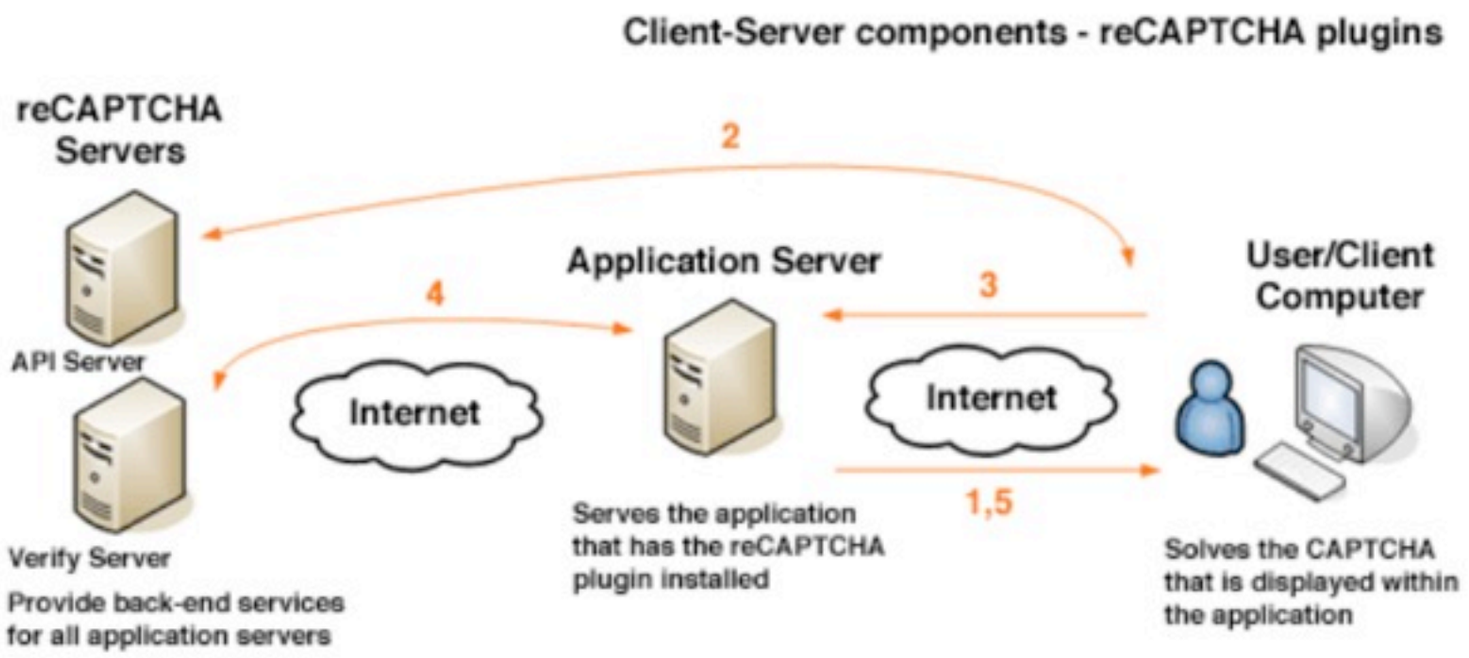
- Example (2): reCAPTCHA



reCAPTCHA

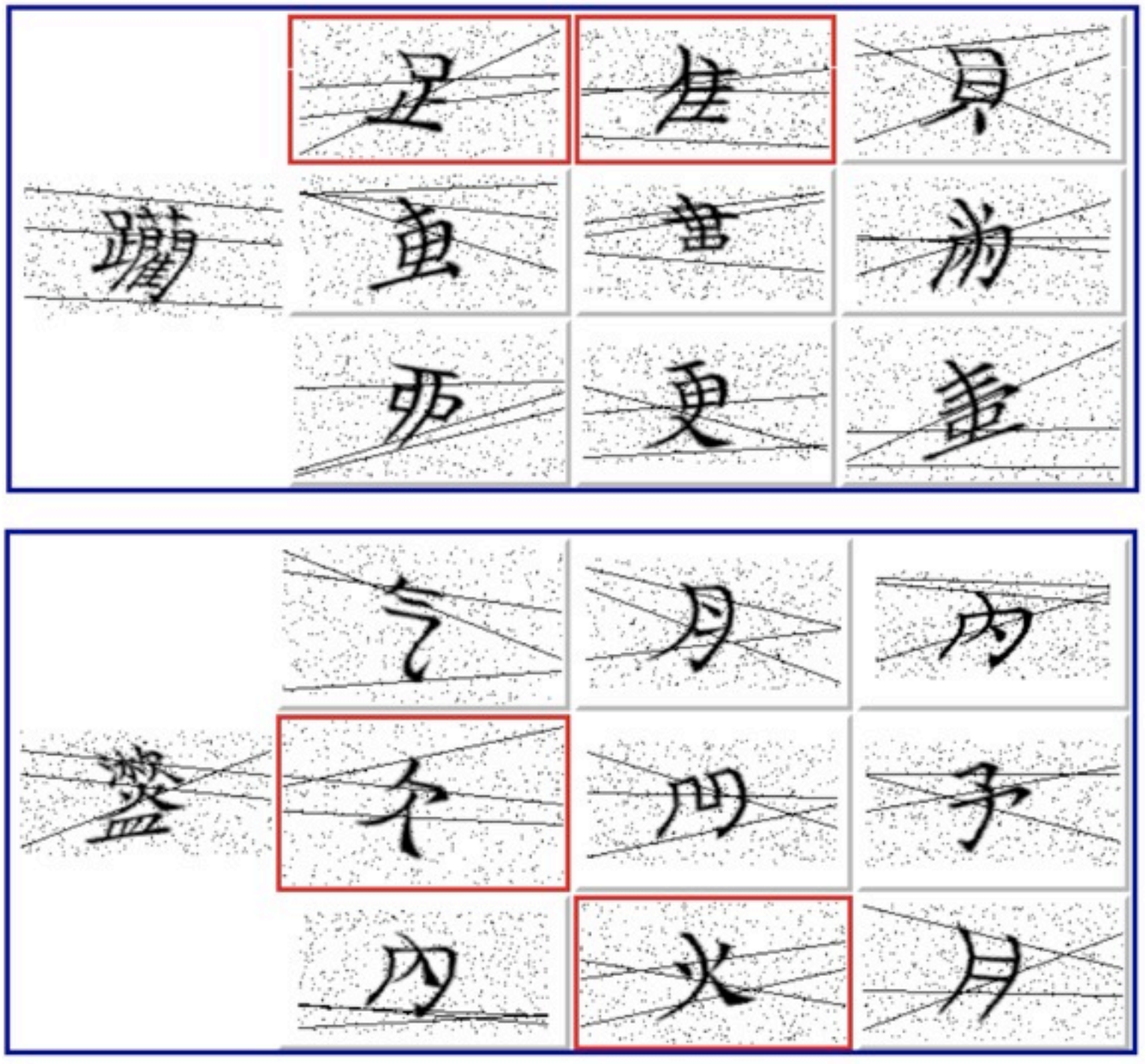



olcott
have
Type the two words:
reCAPTCHA™
stop spam.
read books.
Submit
The words above come from scanned books.
By typing them, you help to digitize old texts.


Client-Server components - reCAPTCHA plugins
reCAPTCHA Servers
API Server
Verify Server
Provide back-end services for all application servers
2
4
Internet
Application Server
Serves the application that has the reCAPTCHA plugin installed
3
Internet
1,5
User/Client Computer
Solves the CAPTCHA that is displayed within the application

Chinese CAPTCHA

Ling-Jyh Chen, Institute of Information Science, Academia Sinica, Taipei, Taiwan

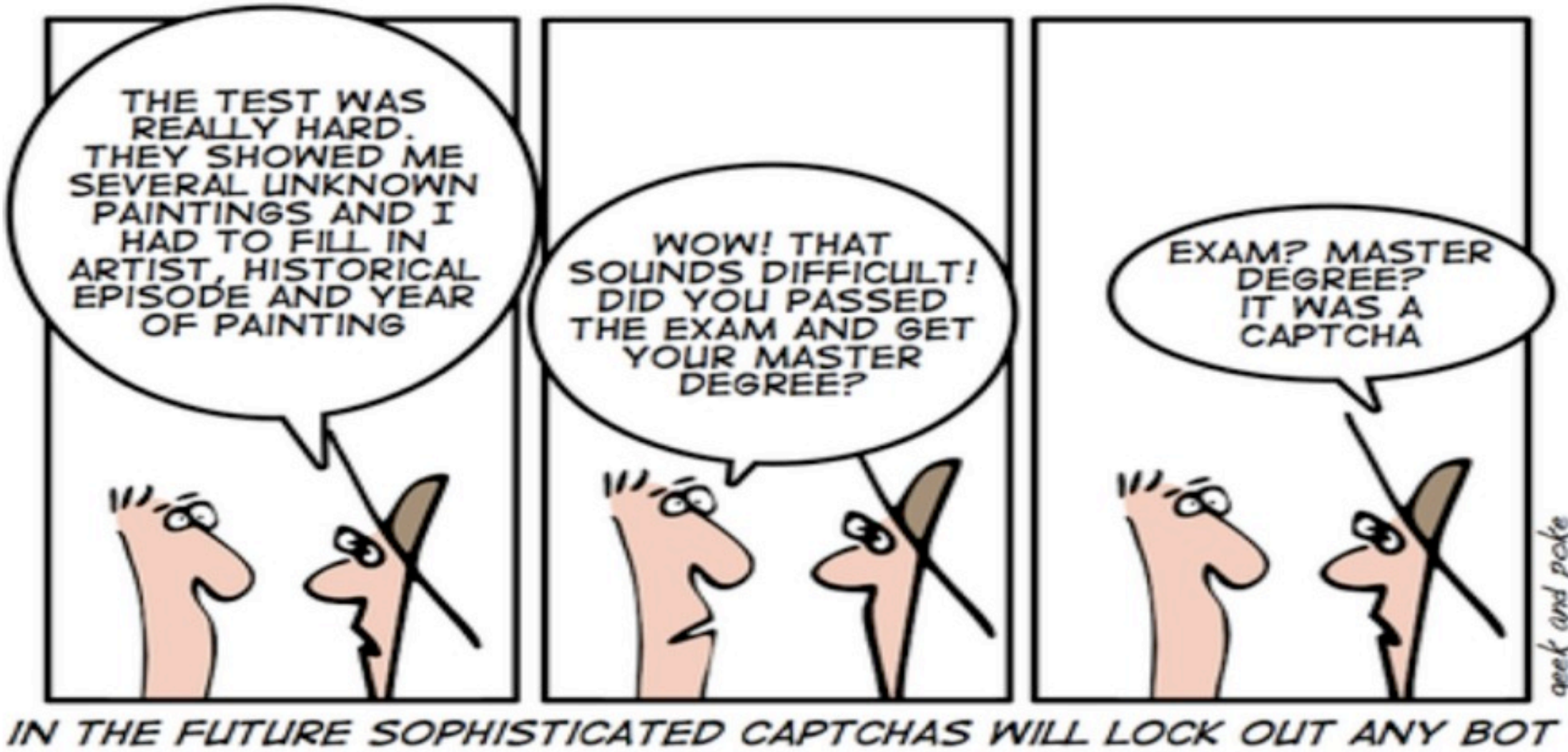


System Design Issues

- Centralized vs. distributed systems
- Single vs. multiple players per round
- Single vs. multiple outcomes per round
- Pure vs. computer-aided HCOMP
- Stationary vs. mobile players
- “just enough” incentives
- Not “just another” HCOMP system

Initiatory Human Computation (4)

- Example (3): KA-CAPTCHA
  - Objective: To collect every correct answer submitted by humans to the CAPTCHA test as a solution to a problem that computers are unable to solve
  - CAPTCHA solvers are highly interested in providing a valid response to the CAPTCHA test (because they want to access the protected resource)
  - Knowledge acquisition mechanism: To strategically asking for a solution to a particular open problem that is of interest to the CAPTCHA designer.

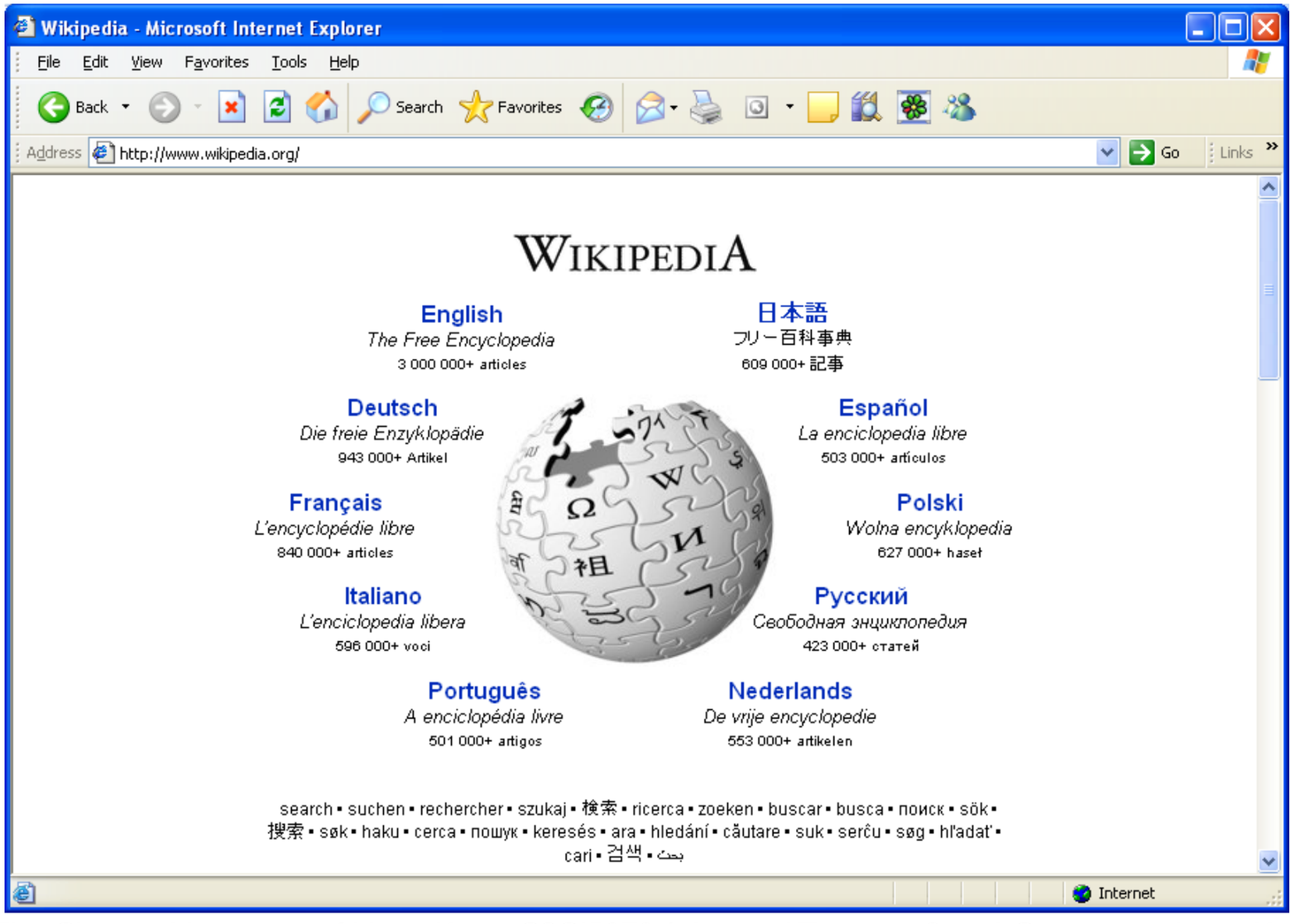

THE TEST WAS REALLY HARD. THEY SHOWED ME SEVERAL UNKNOWN PAINTINGS AND I HAD TO FILL IN ARTIST, HISTORICAL EPISODE AND YEAR OF PAINTING
WOW! THAT SOUNDS DIFFICULT! DID YOU PASSED THE EXAM AND GET YOUR MASTER DEGREE?
EXAM? MASTER DEGREE? IT WAS A CAPTCHA
geek and poke
IN THE FUTURE SOPHISTICATED CAPTCHAS WILL LOCK OUT ANY BOT

Distributed Human Computation (1)

- Objective: To encourage a huge population of Internet users to contribute to solve the difficult AI problems
- Example (1): Razor
  - To use human votes to determine if a given email is spam (anti-spam mechanism)
- Example (2): Proofreader
  - To give a (small) portion of the image file and corresponding text (generated by OCR) side-by-side to a human proofreader

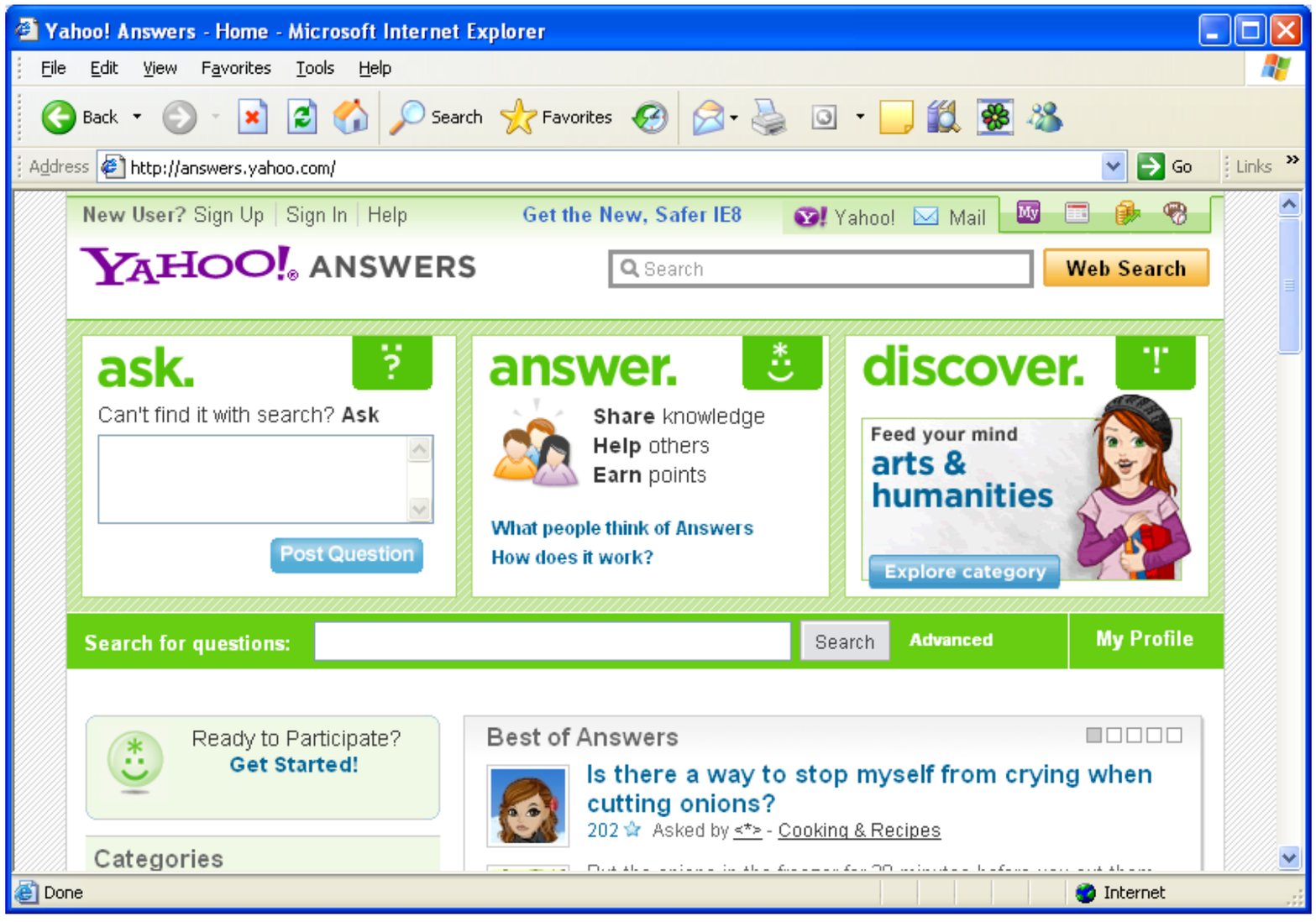
Distributed Human Computation (2)

- Example (3): Wikipedia
  - The collective knowledge is distributed in that essentially almost anyone can contribute to the Wiki



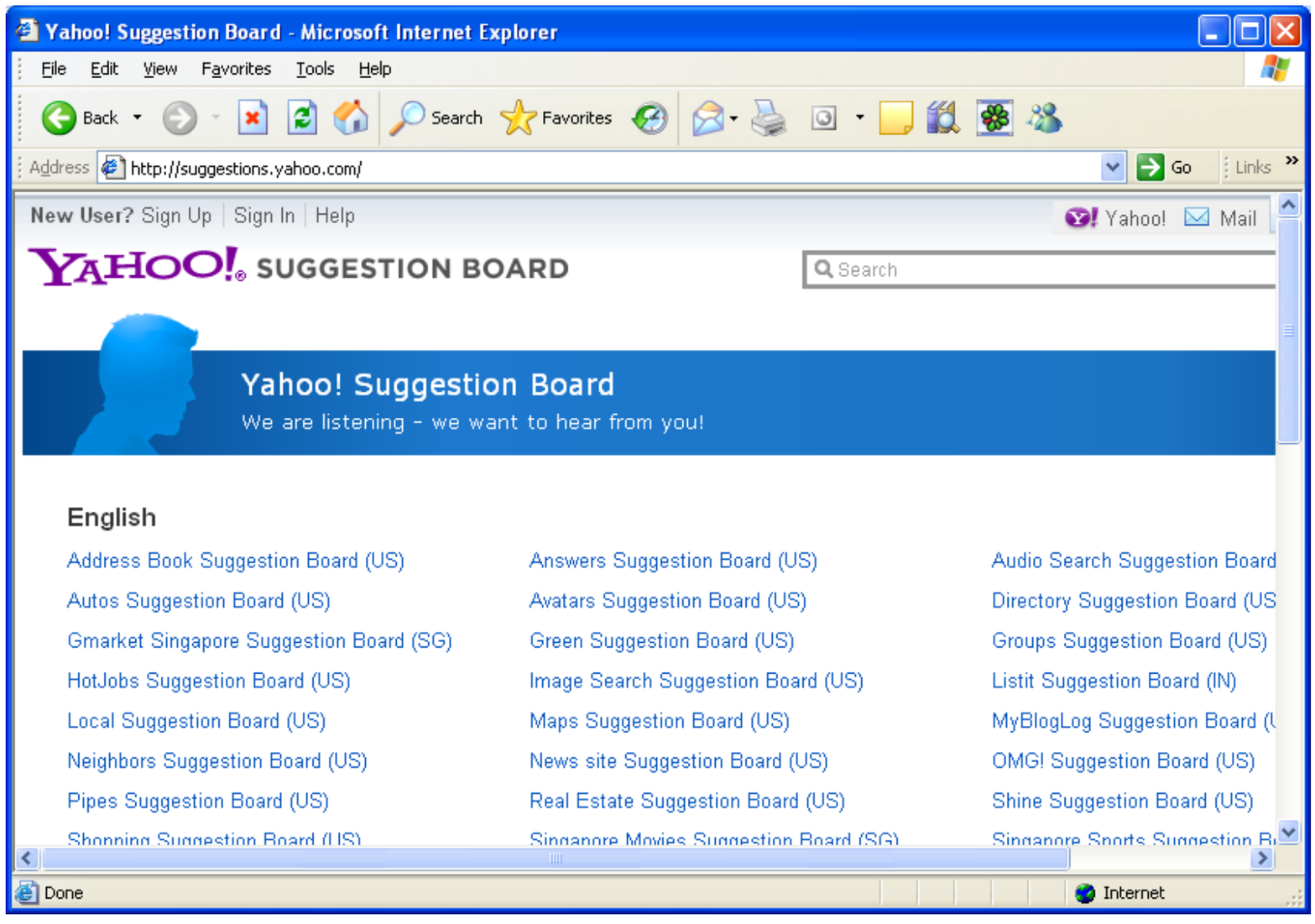
Distributed Human Computation (3)

- Example (4): Yahoo! Answers
  - To provide automated collection of human reviewed data at Internet-scale



Distributed Human Computation (4)

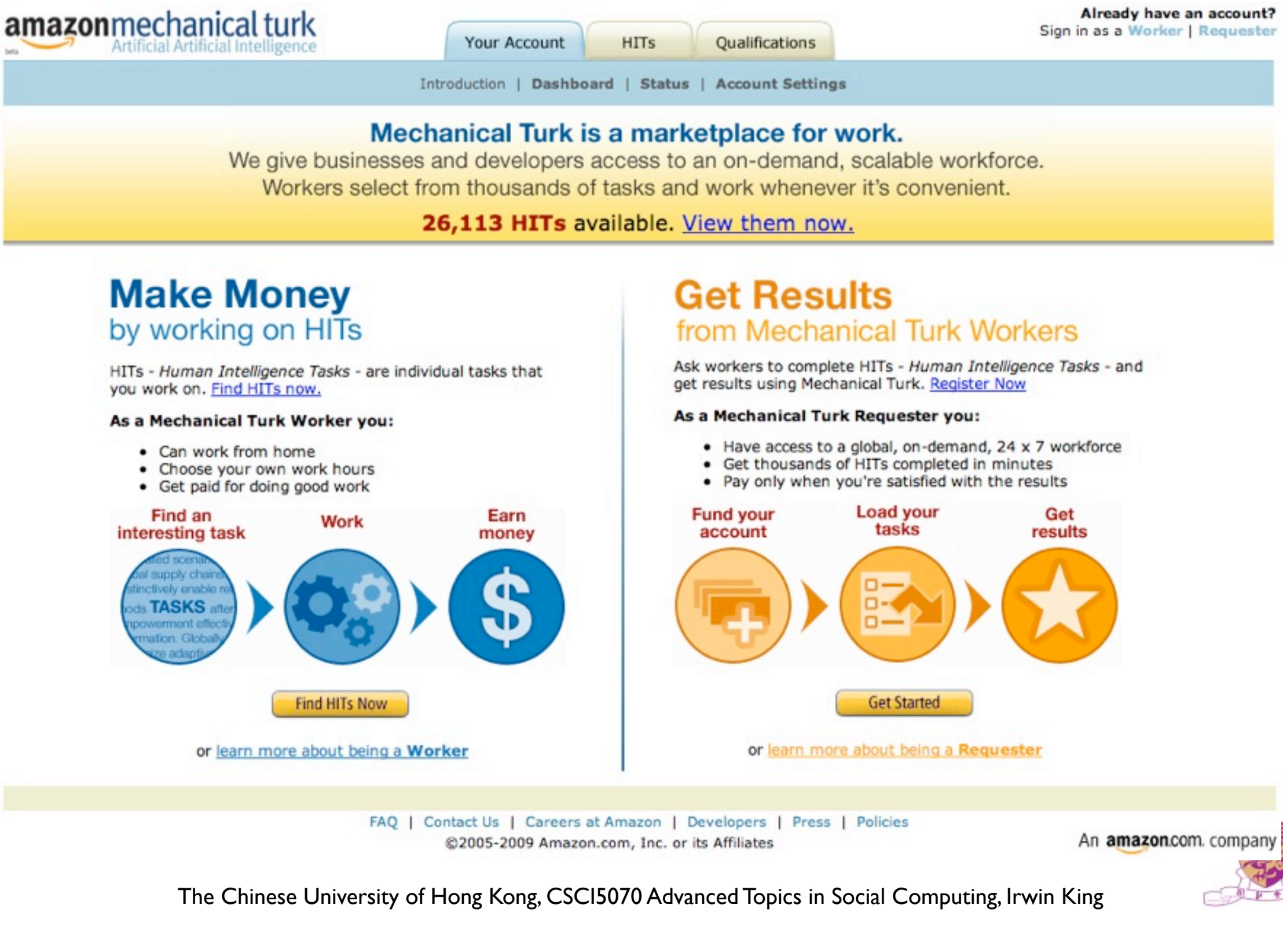
- Example (5): Yahoo! Suggestion Board
  - An Internet-scale feedback and suggestion system



Distributed Human Computation (5)

- Example (6): Amazon Mechanical Turk
  - It provides monetary rewards for tasks
- Example (7): LabelMe
  - A web-based tool for image annotation
  - Anybody can annotate image using it. You can only have access to the database once you have annotated a certain number of images.
- Example (8): 43Things
  - To collect goals from users and help them to find other users who have similar goals
- Example 9: MajorMiner
  - Music annotation game

Amazon Mechanical Turk



Example of Mechanical Turk

Answer a short survey

1. What is your gender?

- ○ Male
- ○ Female

2. What is your age?

3. Which of the following best describes your highest achieved education level?

Some High School

4. What is the total income of your household?

Less than $12,500
$12,500 - $24,999
$25,000 - $37,499
$37,500 - $49,999

5. What is your favorite type of TV Show? (select all that apply)

- ☐ Sports
- ☐ Situational Comedies
- ☐ Drama
- ☐ News
- ☐ Music Videos

Find the Website Address for this Restaurant

- For this restaurant below, enter the website address for the official website of the restaurant
- Include the full address, e.g. http://www.thecheesecakefactory.com
- Do not include URLs to city guides and listings like Citysearch.

Restaurant Name: **${name}**

Address: **${address}**

Phone Number: **${phone}**

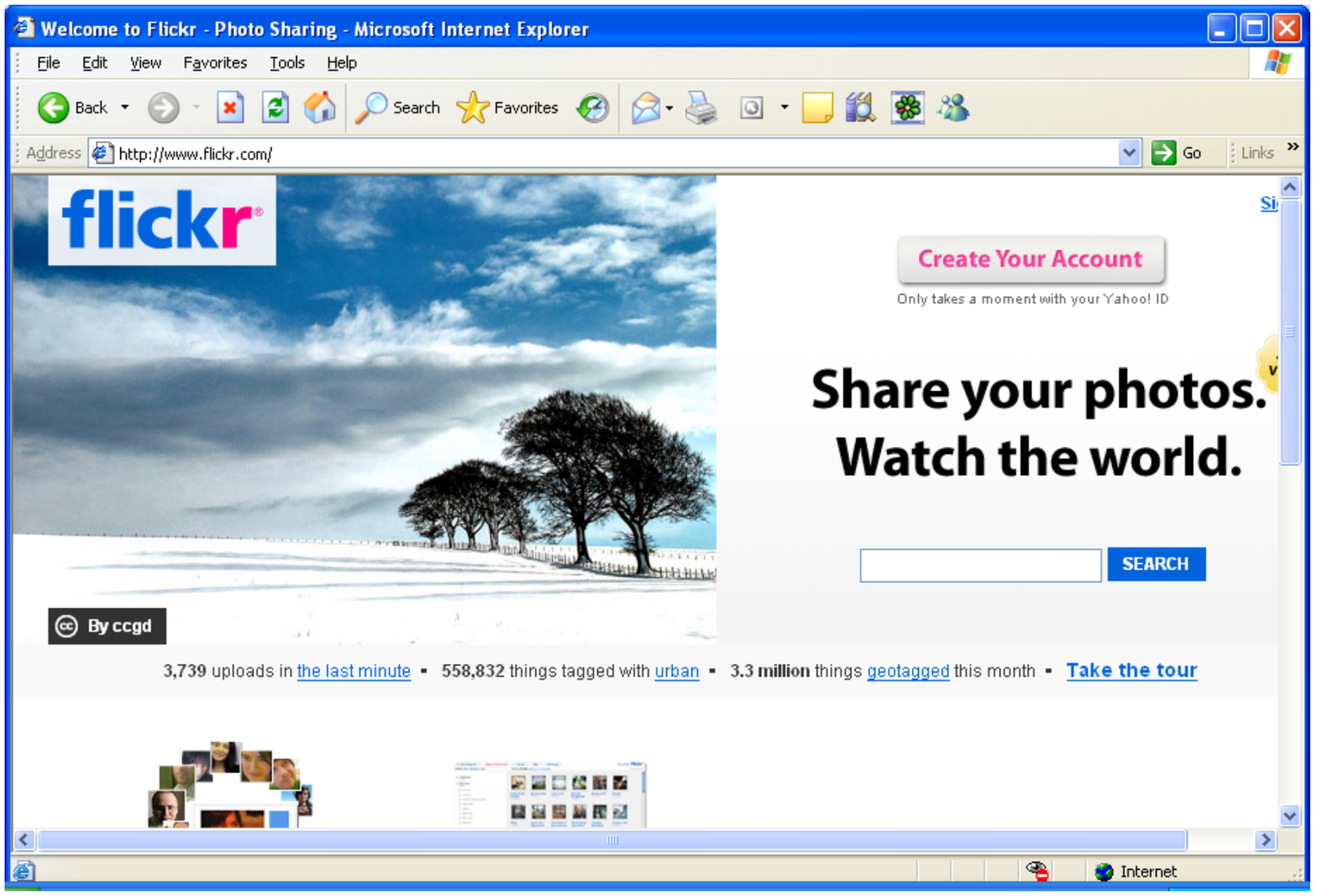
Website:

Please provide any comments you may have below, we appreciate your input!

Submit

Distributed Human Computation (6)

- Example (10): Yahoo's flickr
  - It is a photo-sharing site with captions being used as photo tags



Social Game-based Human Computation with volunteers or paid engineers (1)

- Recently social games were proposed to collect accurate information from players as a side effect of their playing
- The players are volunteers or paid engineers
- Disadvantages:
  - Rely on online volunteers or paid engineers to enter information explicitly
  - Unable to scale up the system due to high cost
  - No validation mechanism to guarantee that the information collected is accurate

Social Game-based Human Computation with volunteers or paid engineers (2)

- Most of the games at early stage aimed to collect commonsense knowledge.
- Example (1): Cyc
  - To collect information from the input by paid knowledge engineers
- Example (2): Open Mind
  - To collect common sense knowledge from people to develop intelligent software
  - Shortcoming: was too reliant on the unpaid volunteers to donate their time to contribute information

Social Game-based Human Computation with volunteers or paid engineers (3)

- Example (2): Open Mind

Social Game-based Human Computation with volunteers or paid engineers (4)

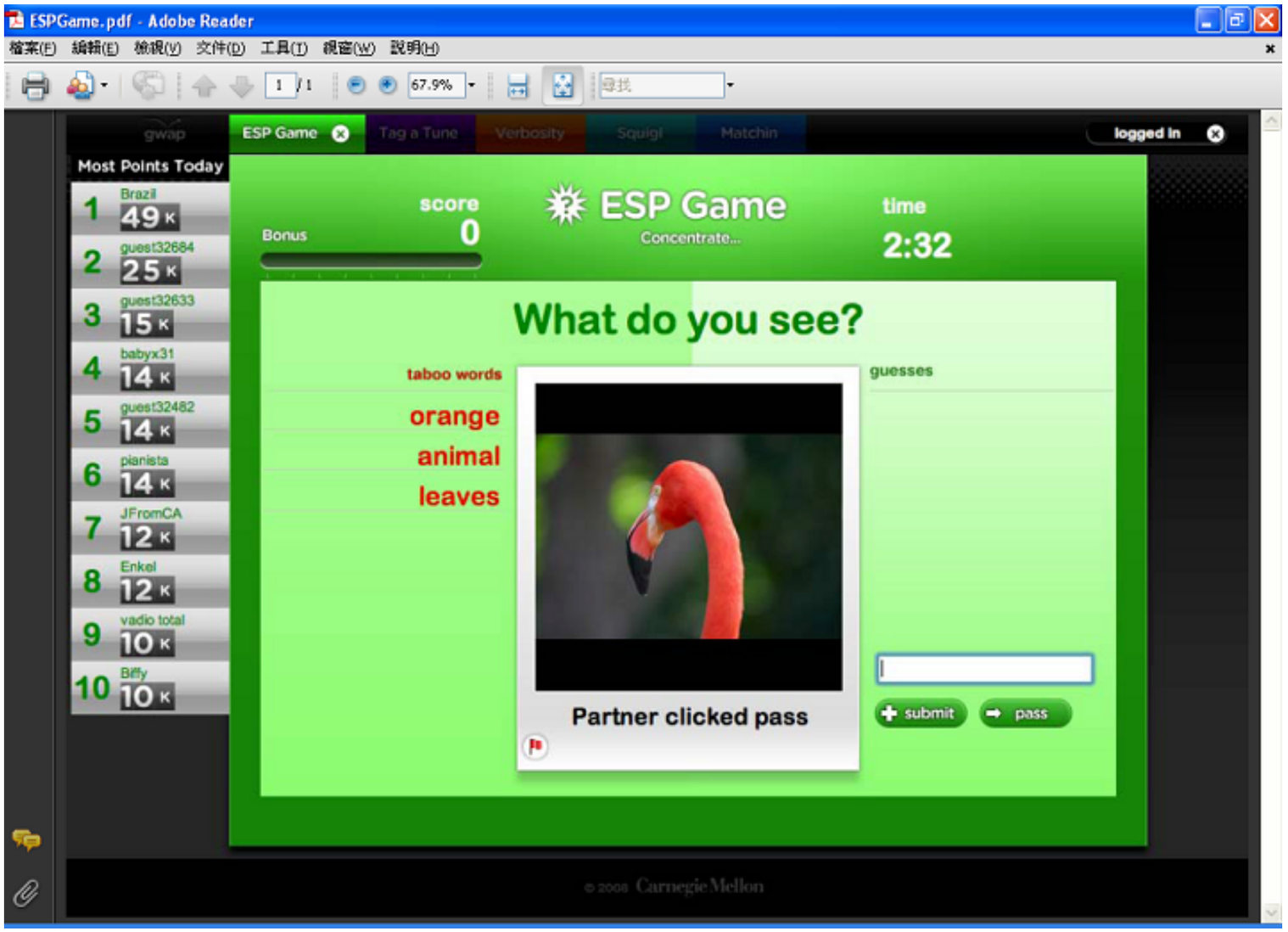
- Example (3): Mindpixel
  - Reward those Internet users who consistently validate a fact inline with the other users
  - Shortcoming: the cost is high!
- Example (4): Wildfire wally
  - To solve the maximum clique problem
  - Shortcoming: rely on unpaid volunteers to donate their time to contribute information

Social Game-based Human Computation with online players (1)

- Later, social games were proposed to collect information from the players as a side effect of their playing
- Advantage:
  - It encouraged more Internet users to contribute information to solve the AI problems because of the increasingly popularity of online game
- TWO important factors for collecting information effectively from players through a social game:
  - Guarantee the quality of collected information
  - Maintain the enjoyment of players in the game

Social Game-based Human Computation with

- To collect text information from images
  - Examples (1): ESP game



Social Game-based Human Computation with

- To collect text information for images:
  - Examples (2): Peekaboom

Social Game-based Human Computation with

- To collect commonsense knowledge:
  - Examples (3): Verbosity

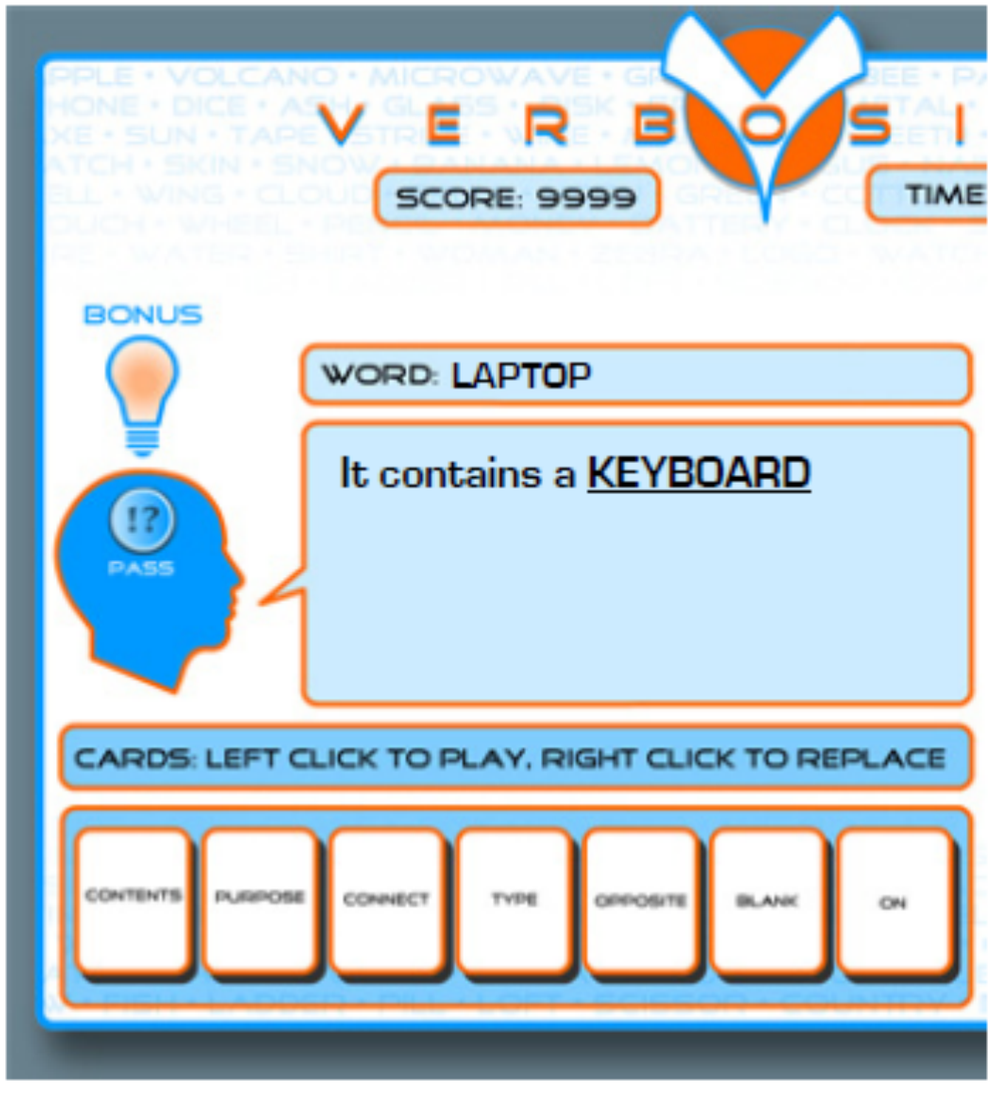
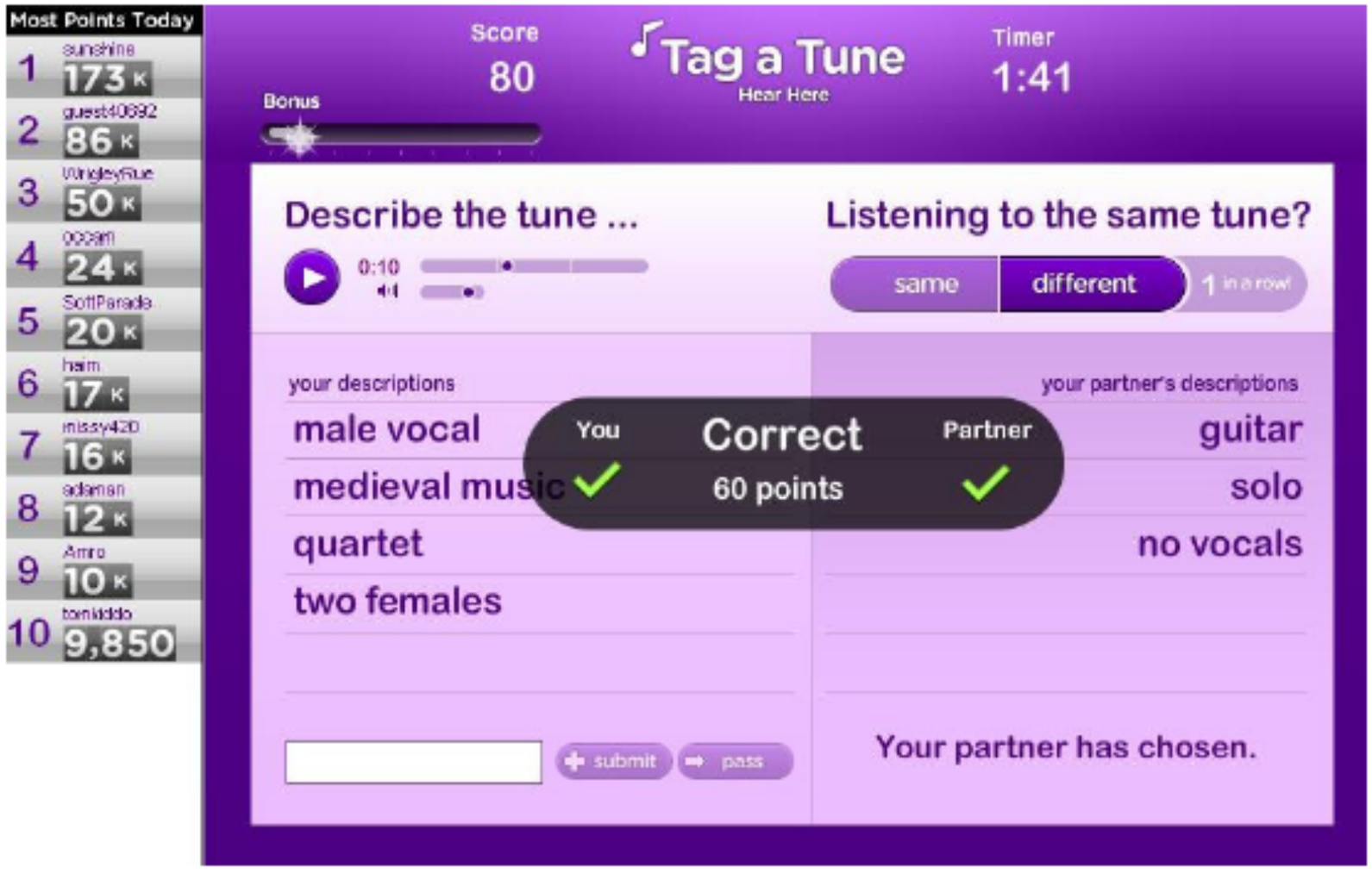


Figure 1. Part of the Narrator's screen.

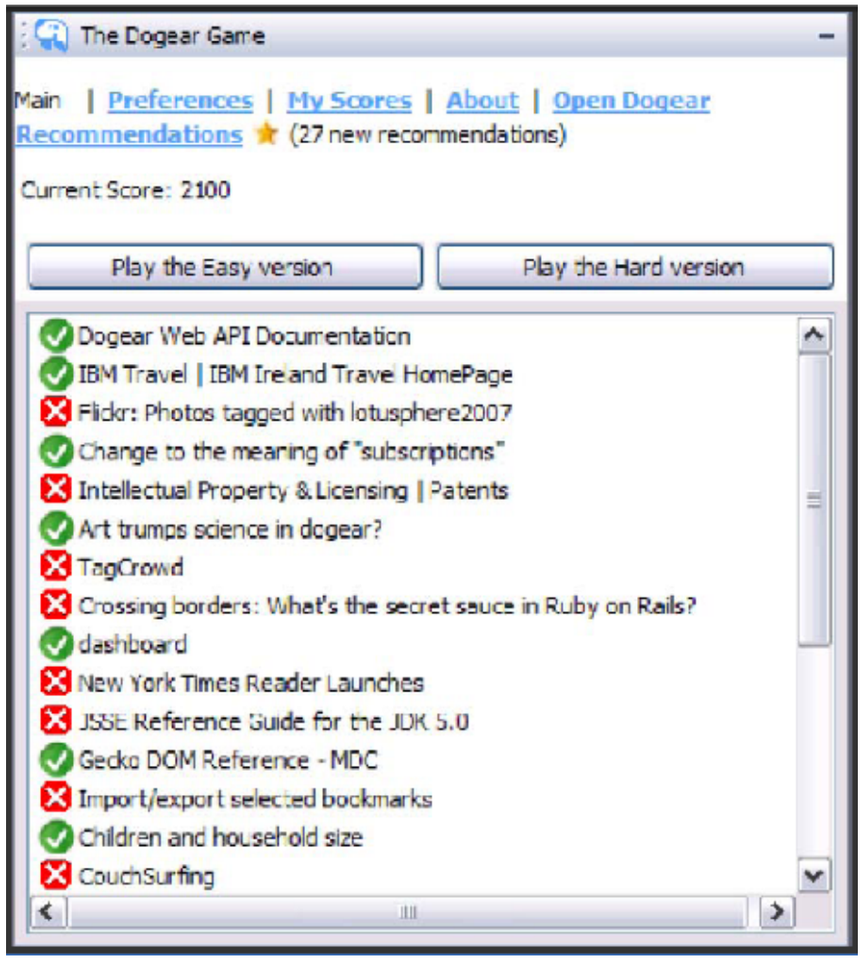
Social Game-based Human Computation with

- To collect subjective descriptions of sounds and music:
  - Example (4): Tagatune



Social Game-based Human Computation with

- To learn colleagues' bookmarks in an organizational goal:
  - Example (5): Dogear Game



Social Game-based Human Computation with

- To tag locations in the real world through gameplay in mobile social games:
  - Example (6): Gopher guessing game

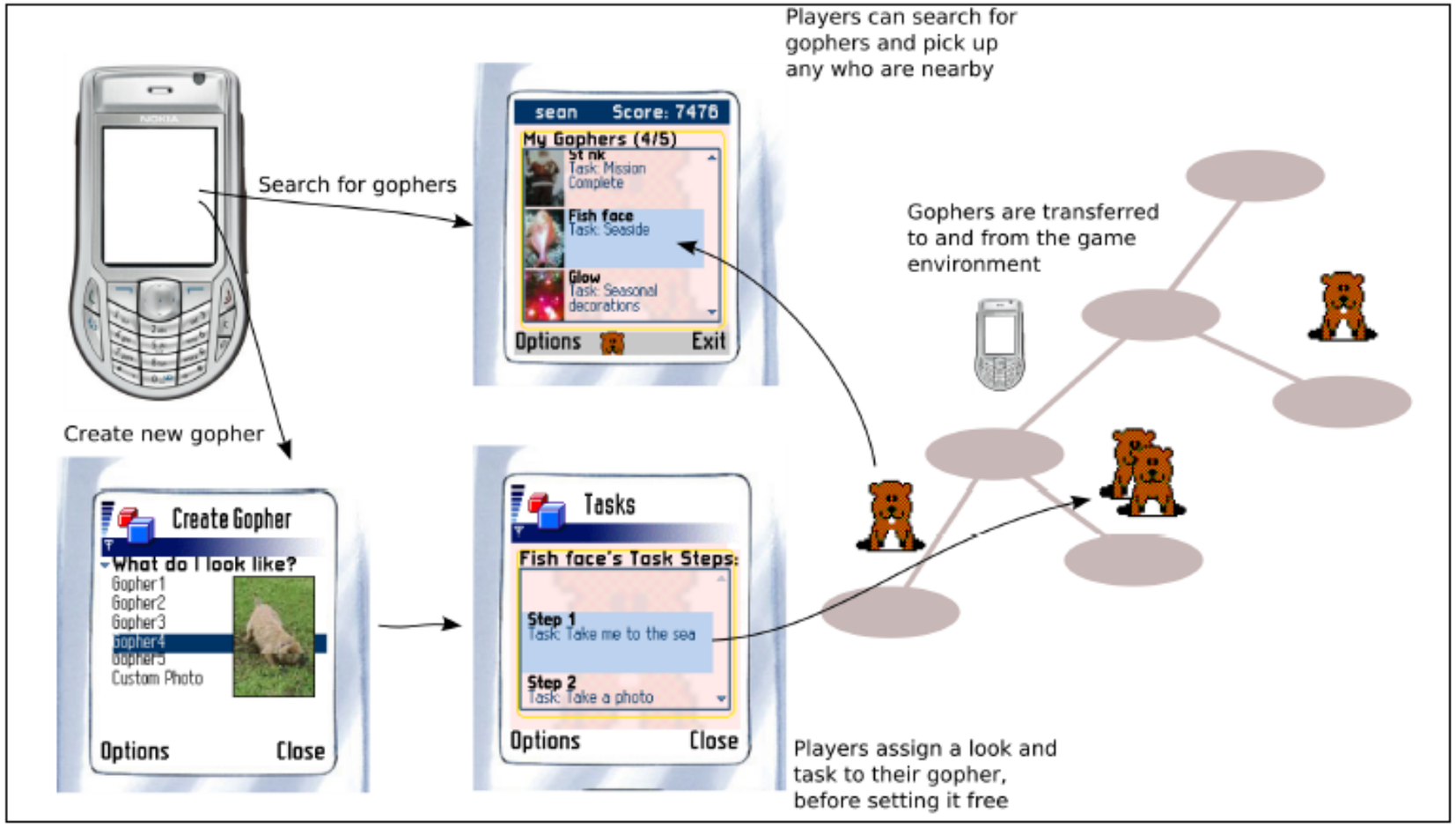


Figure 1. Real world experience, acquiring gophers

Social Game-based Human Computation with

- To tag locations in the real world through gameplay in mobile social games:
  - Example (7): Gopher guessing game

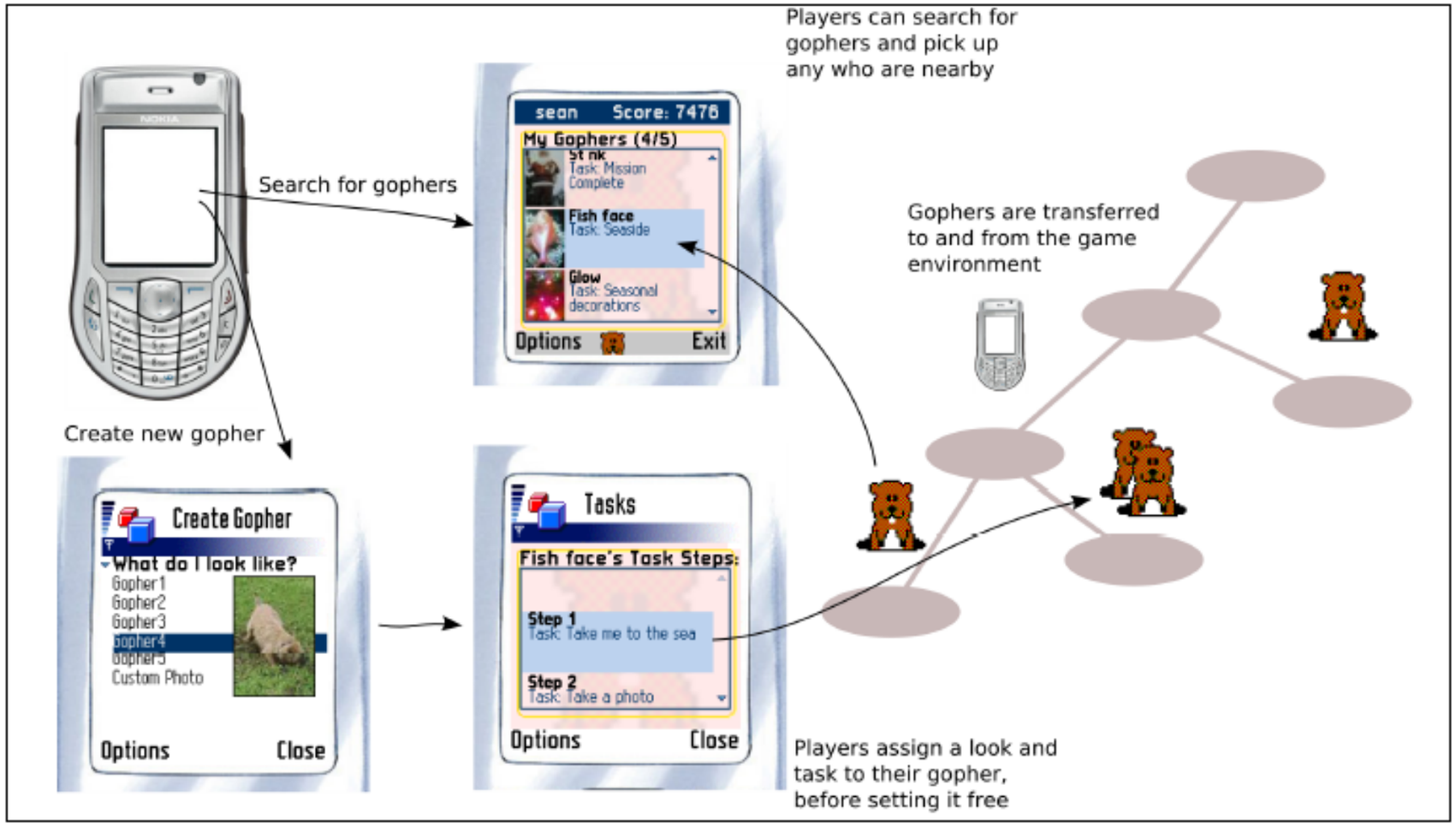
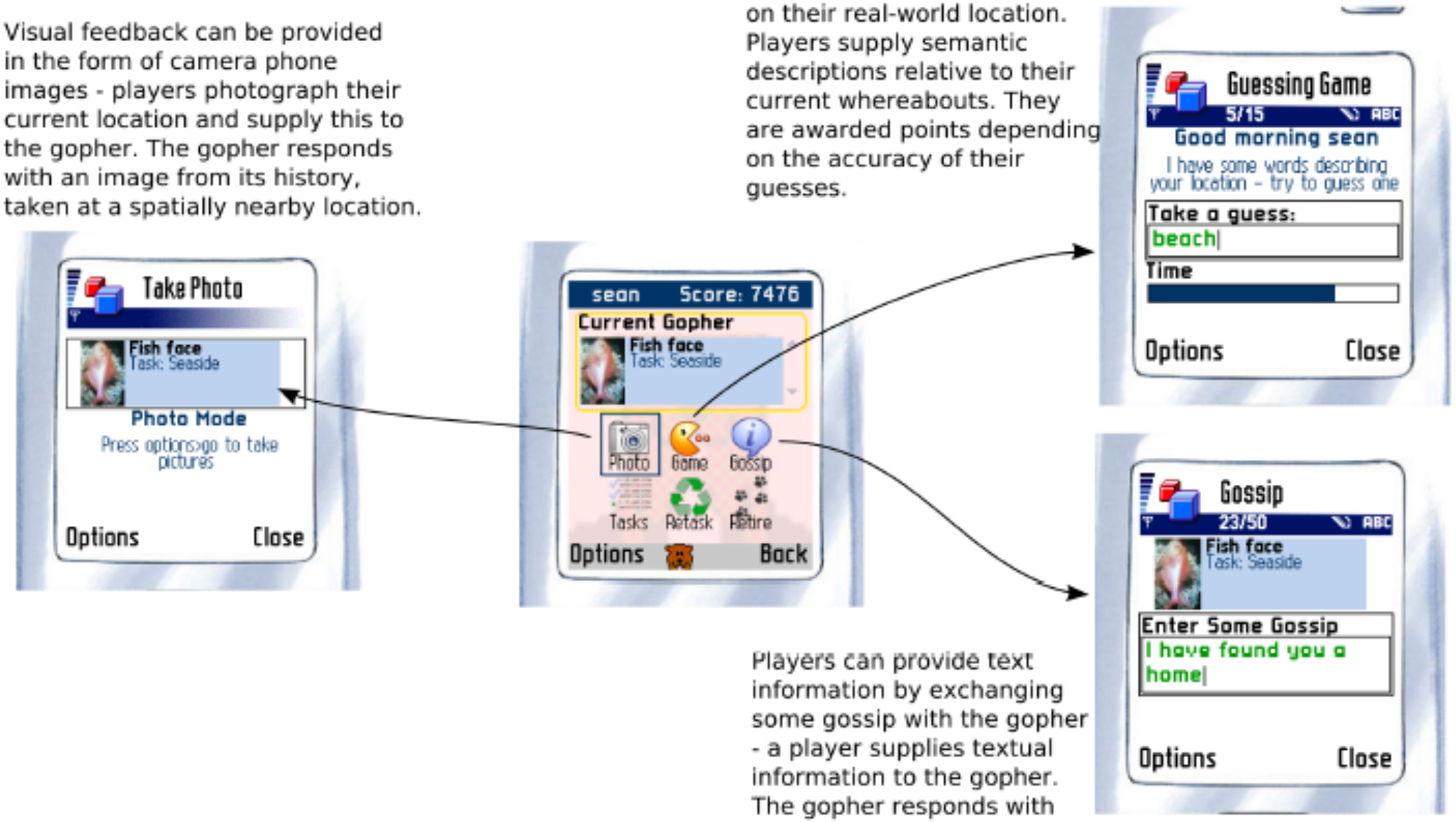


Figure 1. Real world experience, acquiring gophers

Visual feedback can be provided in the form of camera phone images - players photograph their current location and supply this to the gopher. The gopher responds with an image from its history, taken at a spatially nearby location.

Gophers can participate in a word guessing game, based on their real-world location. Players supply semantic descriptions relative to their current whereabouts. They are awarded points depending on the accuracy of their guesses.



Players can provide text information by exchanging some gossip with the gopher - a player supplies textual information to the gopher. The gopher responds with some gossip from it's history, taken at a nearby location.

Figure 2. Real world experience, interacting with gophers

Properties of Social Games

1. Type of information to be collected

2. Game Structure

   1. Output-agreement Game
   2. Input-agreement Game
   3. Inversion-problem Game
   4. Output-optimization Game

3. Verification Method

   1. Symmetric
   2. Asymmetric

4. Game Mechanism

   1. Collaborative
   2. Competitive
   3. Hybrid

5. Player Requirement

Categorization of Social Games

TABLE I
CATEGORIZATION OF SOCIAL GAMES

Game Structure	Verification Method	Game Mechanism
Output-agreement	Symmetric	Collaborative or Hybrid
Input-agreement	Symmetric	Collaborative or Hybrid
Inversion-problem	Asymmetric	Collaborative or Competitive or Hybrid
Output-optimization	Symmetric or Asymmetric	Collaborative or Competitive or Hybrid

Subjective vs. Objective Information

- For subjective information, the information presented for the same subject is affected by users because of different choices of vocabularies for the same subject.
  - lower probability on players' correct outputs being the same
- For objective information, the information presented for the same subject is NOT affected by users because of same choices of vocabularies for the same subject.
  - higher probability on players' correct outputs being the same

Game Structure (1)

- Game structure defines the key elements of a game including players' input, players' output, the relationship among the input and output of players, and the winning condition

- Four types of game structure
  1. Output-agreement Game
  2. Input-agreement Game
  3. Inversion-problem Game
  4. Output-optimization Game

Game Structure (2)

- Output-agreement Games: All players are given the same input and must produce outputs based on the common input
  - An output-agreement game should be used to collect objective information
- Input-agreement Games: All players are given inputs that are known by the game (but not by the players) to be the same or different. The players are instructed to produce outputs describing their input, so their partners are able to assess whether their inputs are the same or different. Players see only each other's outputs
  - An input-agreement game should be used to collect subjective information

Game Structure (3)

- Inversion-problem Games: The first player has access to the whole problem and gives hints to the second player to make a guess. If the second player is able to guess the secret, we assume that the hints given by the first player are correct.

- Output-optimization Games: All players are given the same input and their outputs are the hints of other players' outputs.

  - An output-optimization game should be used to collect subjective information, because the output pattern of players reflects outputs of players are strongly affected by others' outputs. It is subjective.

Verification Methods

- Verification method of a game defines the method to check the output accuracy of players by asking players to do the same task or different tasks
- Symmetric Verification Games: Either an output-agreement game or an input-agreement game is symmetric verification
- Asymmetric Verification Games: Players are assigned to one of the roles to do different tasks

Game Mechanism

- Game mechanism defines the relationship of all players in the game in order to achieve the winning condition
- Collaborative Games determine the winning condition of all players. The accuracy of output is guaranteed by collaboration of all players.
- Competitive Games determine the winning condition of a player. Output accuracy is guaranteed by information stored in a database. Players' enjoyment in the game can be increased in competition.
- Hybrid Game

Player Requirements (1)

- Player requirement defines the rules on accessing the game of all players.
- In Synchronous Games, players have to give real-time response to other players' action.
- In Asynchronous Games, players do not have to give real-time response to other players' action. The information collected from one player is stored in a database and will be used to determine the correctness of other players' output.

Player Requirements (2)

- Number of players define the following types:
- Single-player Games: It allows one player to play and the other's moves can be simulated from the prerecorded game. Only inversion-problem game can be a single-player game.
- Two-player Games: It allows two players to play together.
- Multi-player Games: It allows multiple players to play together. Only hybrid games can be a multi-player game.

Summary

TABLE II
CATEGORIZATION OF SOCIAL GAMES WITH EXAMPLES

Game Structure	Verification Method	Game Mechanism	Player Requirement		Examples
			Num of Player	Game Play	
Output-agreement	Symmetric	Collaborative	2	Synchronous	ESP, Matchi, Squigl, OntoGame
		Hybrid	Multi-players	Synchronous	Common Consensus, Social Heroes
		Hybrid	Multi-players	Asynchronous	Gopher Game
Input-agreement	Symmetric	Collaborative	2	Synchronous	TagATune
		Hybrid	N/A	N/A	N/A
Inversion-problem	Asymmetric	Collaborative	1 or 2	Synchronous	Peekaboom, Verbosity
		Competitive	2	Asynchronous	Dogear, CyPRESS, CARS
		Hybrid	1 or Multi-players	Synchronous	Phetch
Output-optimization	Symmetric	Collaborative	2	Synchronous	Restaurant Game
		Competitive	N/A	N/A	N/A
		Hybrid	Multi-players	Synchronous	Diplomacy

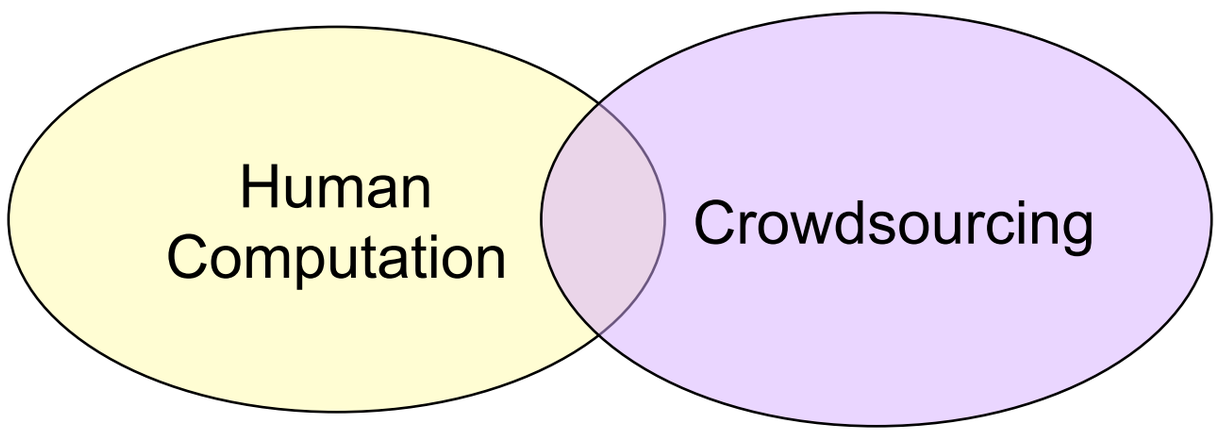
Final Remarks

- Future Work
  - Models, theories, etc.
  - Tools, platforms, etc.
  - Performance metrics, e.g., accuracy, complexity, etc.
- To provide a better understanding about Human Computation Systems (HCS) systematically
- To facilitate future research activities in the field of HCS

Crowdsourcing

Human Computation vs. Crowdsourcing

- Whereas human computation replaces computers with humans, crowdsourcing replaces traditional human workers with members of the public.
- The intersection of crowdsourcing with human computation represents applications that could reasonably be considered as replacements for either traditional human roles or computer roles.



Idea of Crowdsourcing

- To outsource a task that is traditionally performed by an employee to a large group of people (crowd) in the form of an open call

Crowdsourcing

Sheng-Wei (Kuan-Ta) Chen, Institute of Information Science, Academia Sinica, Taipei, Taiwan

- Crowdsourcing = Crowd + Outsourcing
- Soliciting solutions via open calls to large-scale communities
  - INNOCENTIVE
  - oDesk
  - Amazon Mechanical Turk - Marketplace for work
  - Yahoo! Answers
  - Wikipedia

What Are Crowdsourceable?

- Software development - USD $25,000 per job
- Data entry - USD $4.4 per hour
- Image tagging - USD $0.04 per image
- General questions - points on Yahoo! Answers
- Image understanding - USD $0.01 to $0.02 per task
- Human action recognition - USD $0.01 per task
- Linguistic annotations (word similarity) - USD $0.2 per 30 word pairs

Why crowdsourcing is important? (1)

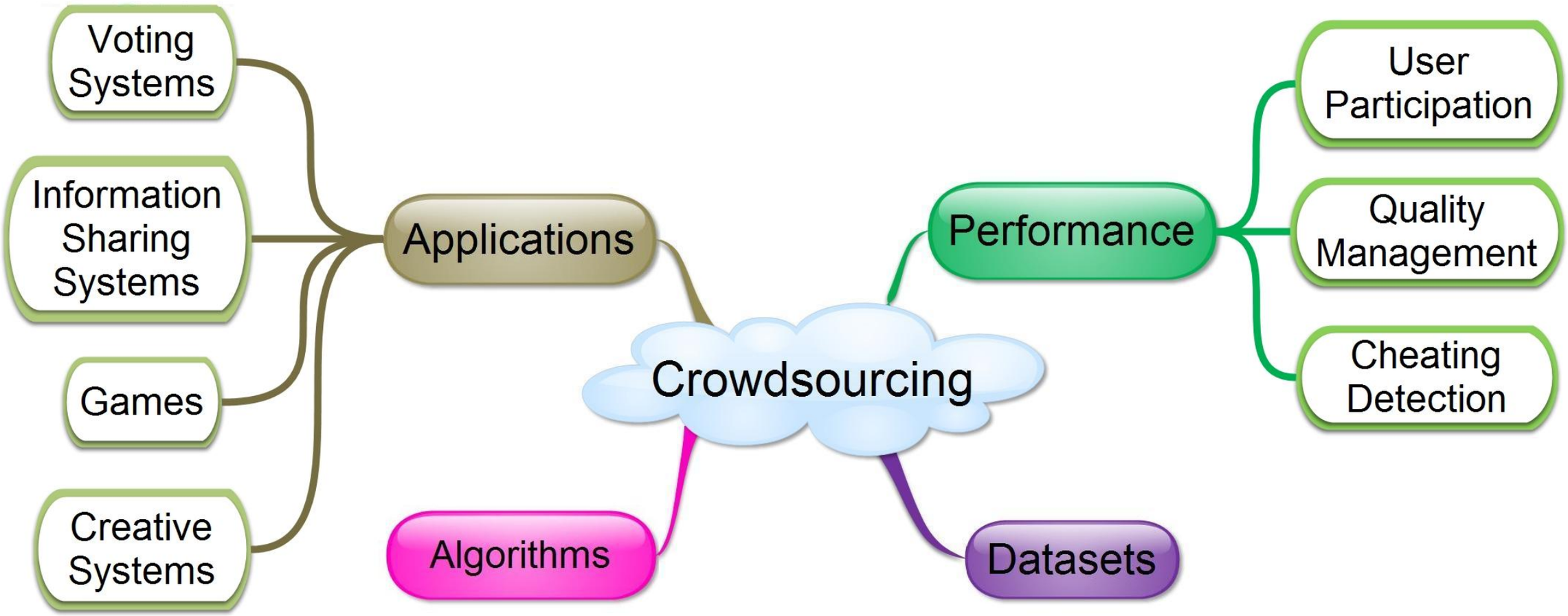
- AMT - Statistics [Ross et al. 2010, Ipeirotis 2010]
- Data available at: http://www.mturk-tracker.com

Collection Period	01/2009 to 04/2010 (16 months)
Number of registered workers	over 400,000
Number of tasks	6,701,406
Number of requesters	9,436
Total value of the posted tasks	US$ 529,259

Why crowdsourcing is important? (2)

- Idea behind
  - Solve some problems which are difficult to be solved by computers
  - Outsource these problem-solving tasks to the crowd rather than an employee in a company
  - Reduce a company's production costs
  - Make more efficient use of labor and resources

Taxonomy of crowdsourcing

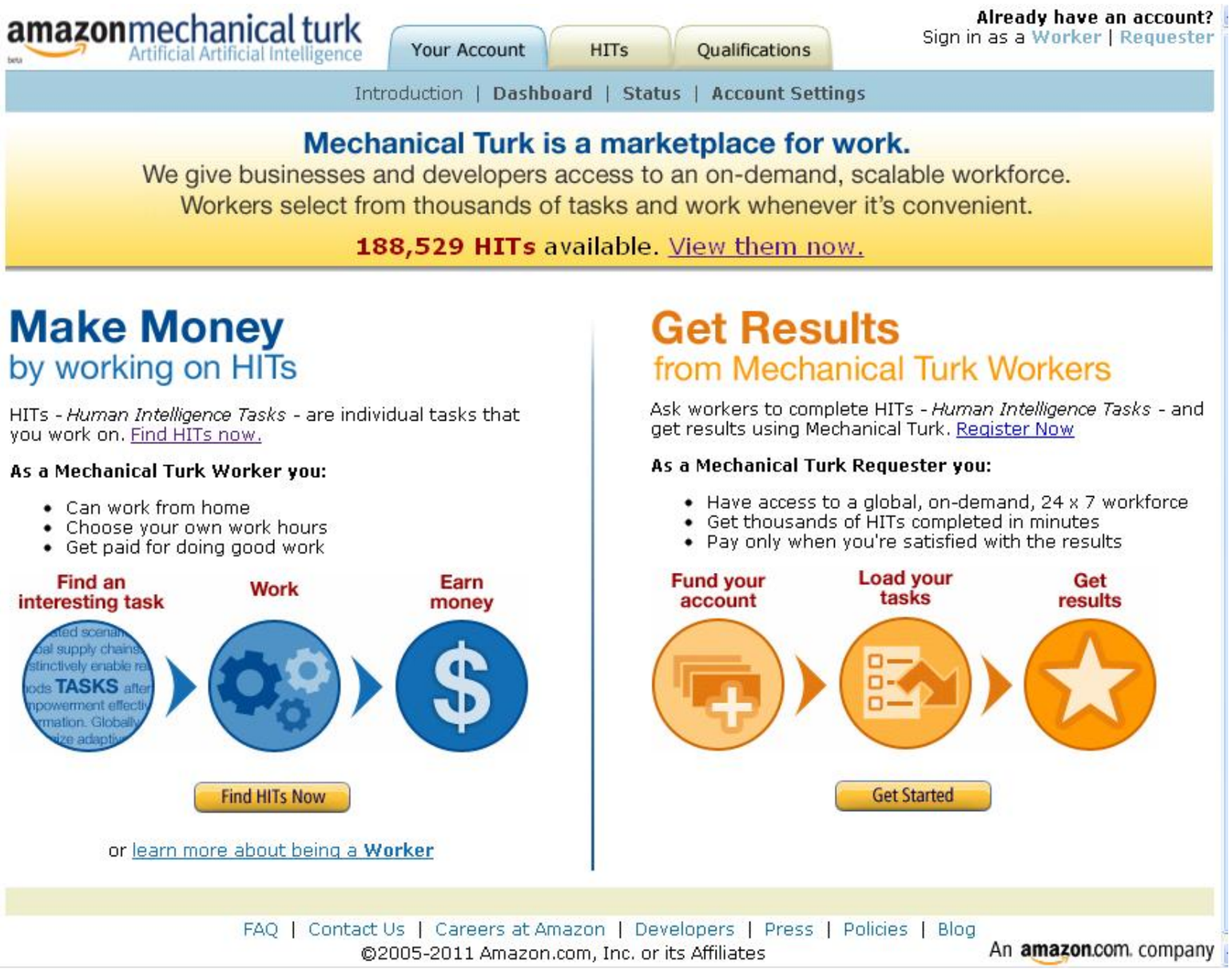


Category 1: Applications

- The categories of applications are:
  - Voting systems
  - Information sharing systems
  - Games
  - Creative systems

Voting systems (1)

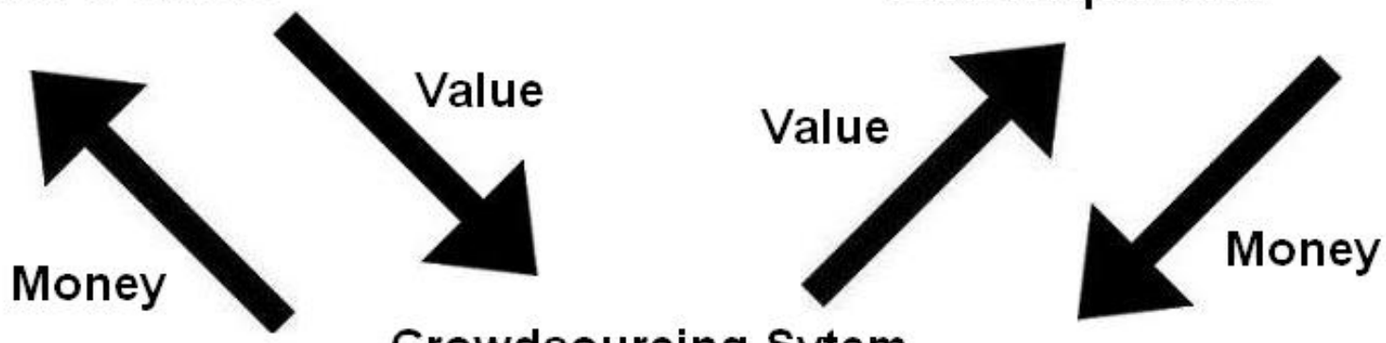
- A large number of applications or experiments were conducted in Amazon Mechanical Turk (AMT) site.



Amazon Mechanical Turk (AMT)

Task Workers

Task Requesters



Make Money
by working on tasks
for requesters

Get Results
from workers by
offering money

Voting systems (2)

- Find out the answer that the majority selected
  - Voting can be used as a tool to evaluate the correctness of an answer from the crowd
- Example (1): Named entity annotation
  - To identify and categorize textual references to objects in the world, such as persons and organizations
- Example (2): Opinions
  - Gathering subjective preferences from the crowd

Voting systems (3)

- Example (3): Geometric reasoning tasks
  - The ability to interpret and reason about shapes is a specific human capability.

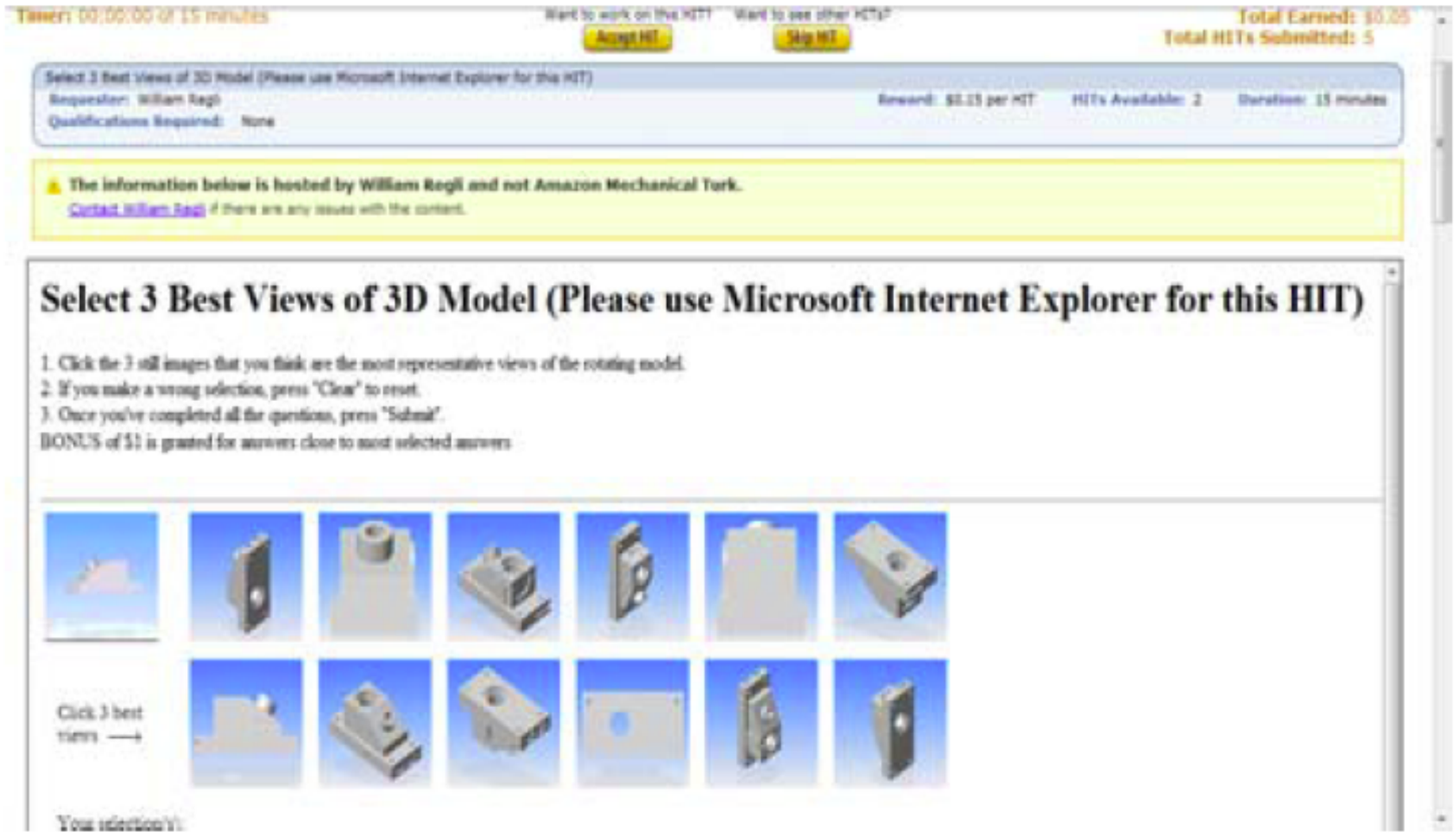


Figure 1. The Canonical Viewpoint HIT

Voting systems (4)

- Example (4): Commonsense
  - Humans can poss commonsense knowledge about the world, but computer programs cannot.
- Example (5): Relevance evaluation
  - Humans have to read through every document in a corpus to determine its relevance to a set of test queries.
  - Each crowdsourcing work perform a small evaluation task.

Voting systems (5)

- Example (6): Natural language annotation
  - Crowdsourcing is a cheap and quick alternative to expert annotations
- Example (7): Spam identification
  - Junk email cannot be determined without the task of understanding content by humans

Information Sharing Systems

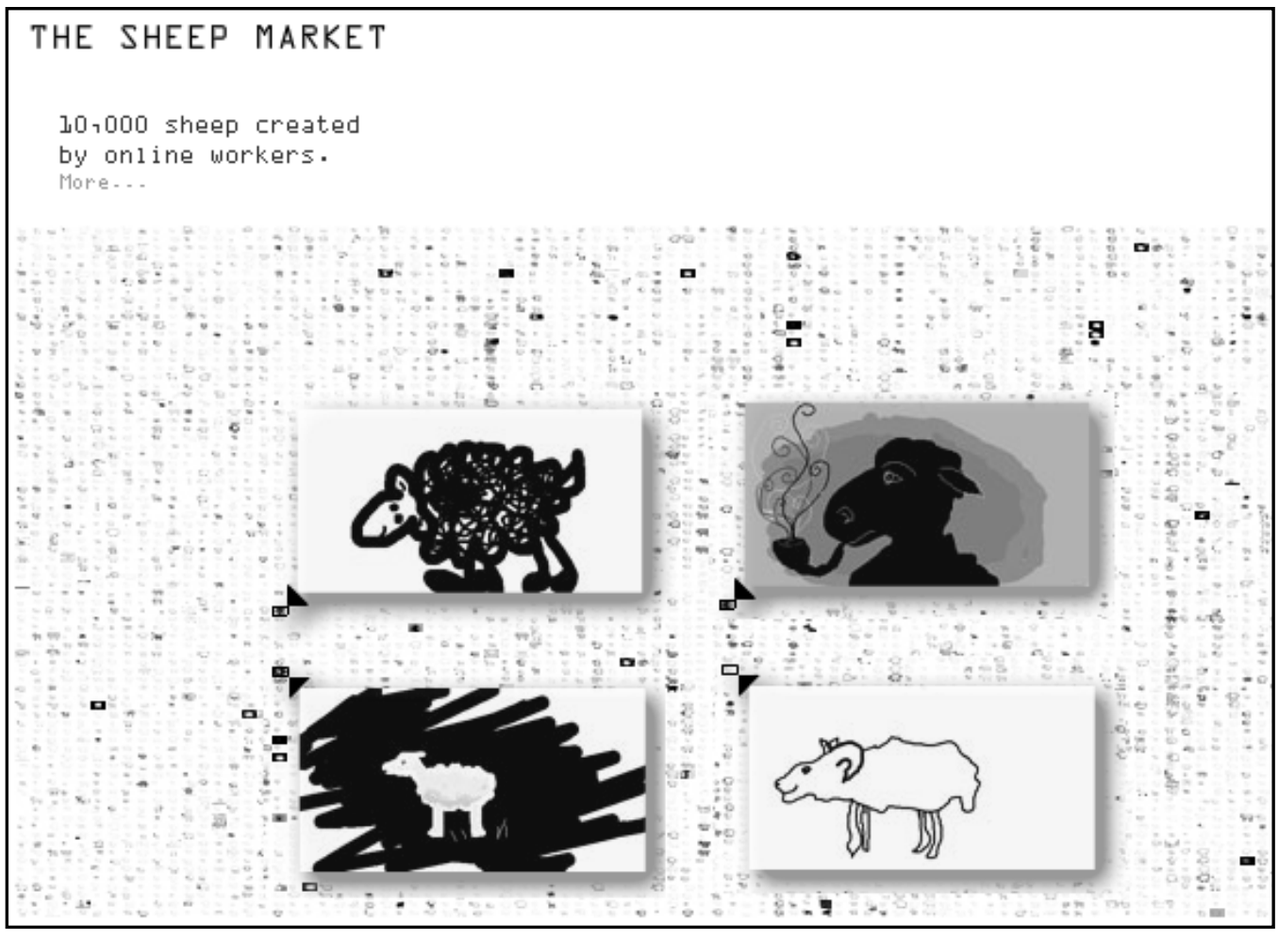
- Share user-generated meta-data among the crowd
- Example (1): Wikipedia
- Example (2): Yahoo! Answer
- Example (3): Yahoo! Suggestion Board
- Example (4): 43Things
- Example (5): Yahoo's flickr
- Example (6): del.icio.us

Games

- Produce useful metadata as a by-product
- Example (1): ESP game
- Example (2): Peekaboom
- Example (3): Verbosity
- Example (4): Tagatune
- Example (5): Gopher guessing game

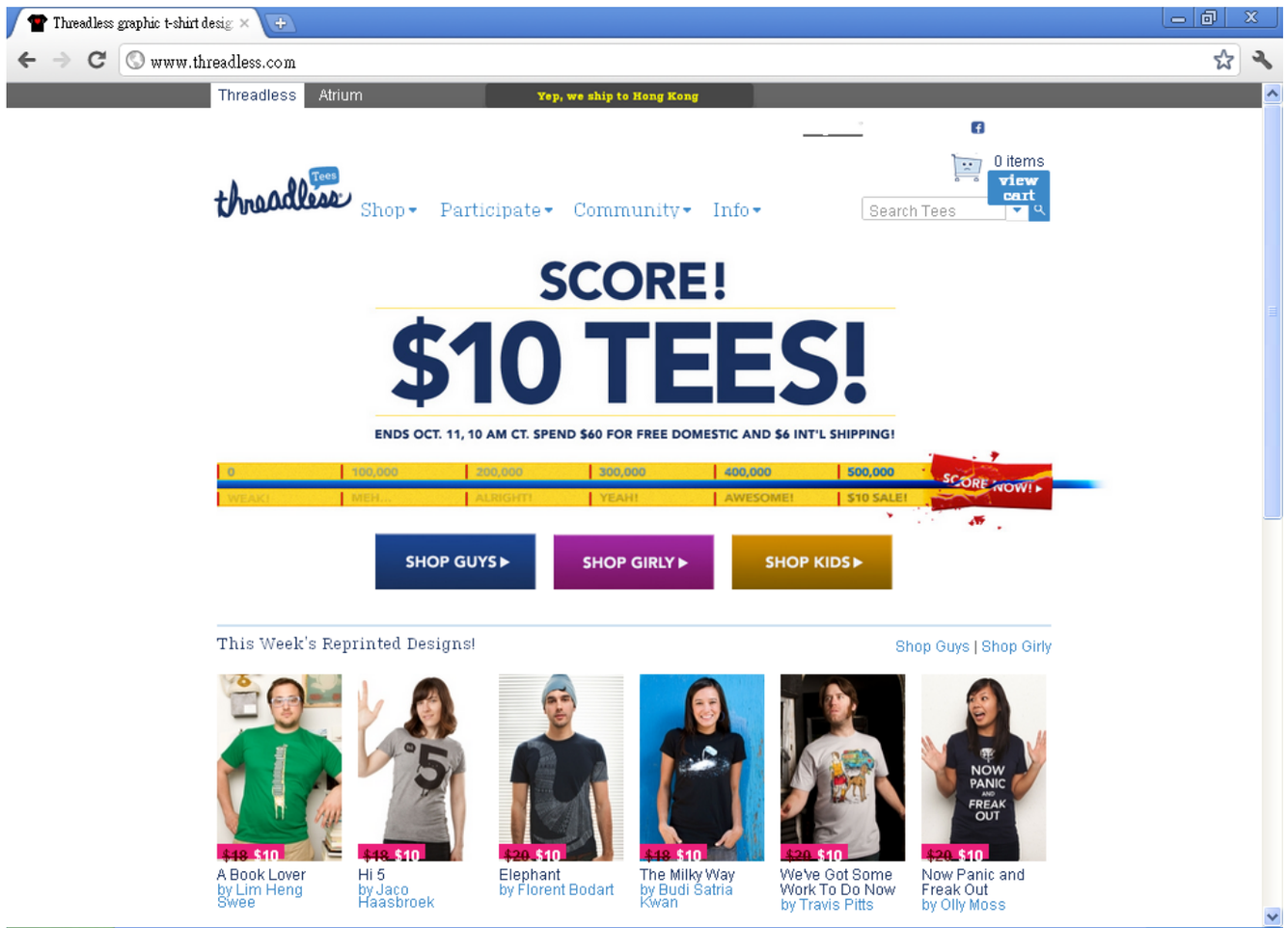
Creative Systems (1)

- Cannot replace the role of human in creativity
- Example (1): The Sheep Market
  - Collect drawings of "a sheep facing to the left" for payment of two cents



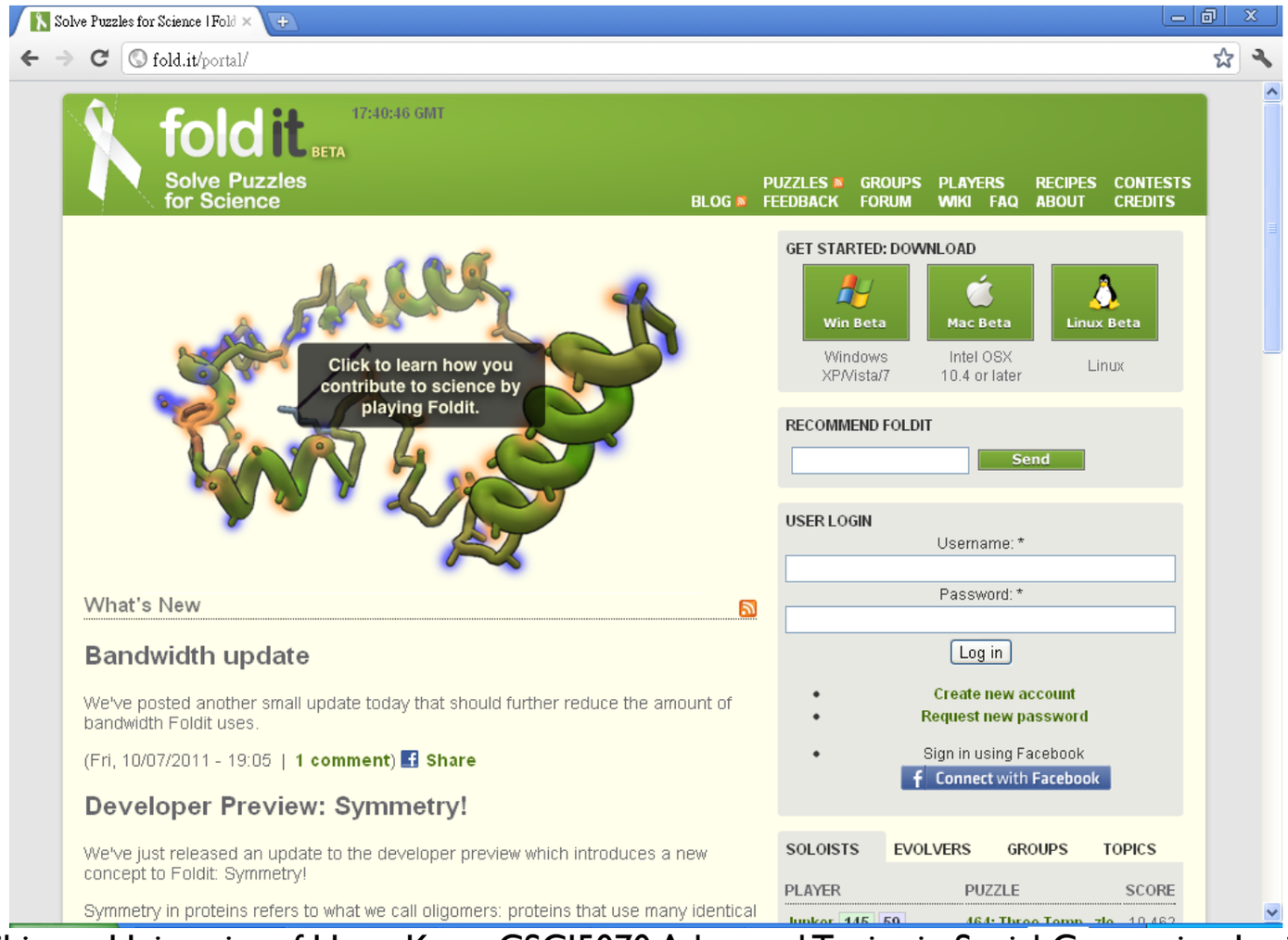
Creative Systems (2)

- Example (2): Threadless
  - To collecting graphic T-shirt designs created by the community
  - Different individuals may create different ideas



Creative Systems (3)

- Example (3): Foldit
- To allow game players to assist in predicting protein structures to find cures for diseases by taking advantage of humans' puzzle-solving intuitions



Category 2: Algorithms

- To make the crowdsourcing systems more efficient and effective to perform their tasks
- To predict the expected time for task completion
- To estimate average precision
- To study the impact of participation rate against the offered rewards
- To analyze user behavior
- To examine motivation patterns

Category 3: Performance

- The categories of performance are:
  - User Participation
  - Quality Management
  - Cheating Detection

User Participation

- Crowd = A population of anonymous Internet users
- Understanding the demographics of crowdsourcing workers, exploring the crowdsourcing incentives and examining their behaviors attracted significant attentions.

Demographics

- The literature showed that:
  - the worker population shifts from a primarily moderate-income, U.S.-based workforce towards an increasingly international group with a significant population of young, well-educated Indian workers
  - they worked as a part- or full-time job
  - professionals, students, and non-workers seem to be more likely to take the task seriously than financial workers, hourly workers, and other workers
  - men over 30 and women of any age were much more likely to qualify

Financial Incentives

- The importance of money compared with other motivations
  - 25 percent of Indian respondents and 13 percent of U.S. respondents reported that MTurk is their primary source of income
- Financial incentives actually encourage quality if the task is designed appropriately
- Q&A sites should function better (faster answers by filling faster the FAQ lists) with both long-term and short-term rewards

Intrinsic Incentives (1)

- Some systems do not offer monetary rewards to their workers. What are the motivations of contribution in these systems?
- Example (1): YouTube
  - Attention, measured by the number of downloads
    - Contributors who stop receiving attention tend to stop contributing
    - Prolific contributors attract an ever increasing number of followers and their attention in a feedback loop
- Example (2): TopCoder.com
  - Highly rated contestants face tougher competition from their opponents in the competition phase of the contest

Intrinsic Incentives (2)

- Example (3): Question-answering sites
  - Altruism, learning, and competency are frequent motivations for top answerers to participate
- Example (4): Open bug reporting
  - In the case of Mozilla, what Mozilla gained was a small pool of talented developers and a number of critical fixes before the release of Firefox 1.0.
  - They have no intention of becoming regular contributors
  - They just want a bug fixed or a feature implemented

Worker Behavior (1)

- User interfaces can affect the behavior of crowdsourcing workers.
  - On AMT, workers are limited by the current user interface and complete tasks by picking the tasks available through one of the existing sorting criteria.
- Many of the same workers completed tasks in multiple batches of tasks, such as document relevant assessment on AMT.
  - However, MTurk cannot prevent this happens.

Worker Behavior (2)

- On Taskcn.com, it found that
  - Most workers become inactive after only a few submissions, while others keep attempting tasks.
- They tend to select tasks where they are competing against fewer opponents to increase their chances of winning.
- They tend to select tasks with higher expected rewards.

Quality Management

- Image Annotation
- Text Annotation
- General Tasks

Image Annotation (1)

- Compare the quality of non-expert annotations and existing gold standard labels for natural language tasks provided by expert labelers.

- It is required to collect an average of 4 non-expert labels per item in order to emulate expert-level label quality, and that the annotation quality can be improved significantly after applying bias correction techniques.

Image Annotation (2)

- With repeated labeling, it is possible to improve the data quality at low cost, especially when labels are noisy

- Different annotators judge the same data and the inter-annotator agreement among different annotators can ensure the quality

Text Annotation

- The use of a qualification test provides the highest improvement of quality of linguistic data collected in MTurk.
- The label quality is affected by cognitive awareness of human knowledge.
- The interannotator agreement can be used as a quality measure for MTurk labels.

General Tasks

- Some works constructed models for predicting the rate and quality of work.
- These models were trained on worker outputs over a set of designs, and were then used to optimize a task's design.
- Some works focused on decomposing the data generation task in a flexible, reusable way. For example:
  - A two-phase, hybrid model for named entity recognition.
  - In the first phase, a trained annotator labels all named entities in a text irrespective of type.
  - In the second phase, naive crowdsourcing workers complete binary judgment tasks to indicate the type(s) of each entity.

Cheating Detection (1)

- Malicious workers often try to maximize their financial gains by producing generic answers rather than actually working on the task.
- Examples:
  - Task-dependent evaluation
  - Interface-dependent evaluation
  - Audience-dependent evaluation

Cheating Detection (2)

- Technique (1): Based on control questions which are evaluated automatically
- Technique (2): Rely on manual checking by the requester
- Technique (3): Hiring experts for fraud detection is very expensive
- Example: to improve precision and recall of current fraud detection techniques for online auction sites

Cheating Detection (3)

- Two crowd-based approaches are proposed recently to detect cheating workers: a majority decision (MD) and an approach using a control group (CG) to re-checking the main task.
- For MD, the same task is given to several different workers and the results are compared. The result which most of the workers submitted is assumed to be correct.
- For CG, a single worker works on a main task and a control group consisting of certain other workers re-checks the result, whether it is valid or not. Usually the main task is expensive, while the re-check task is cheaper. A task is considered to be valid, if the majority of the control group decides the task is correctly done.

Category 4: Datasets (1)

- Collected labels for images, sound clips
  - Example (1): a list of 100,000 images with English labels from the ESP Game
  - Example (2): human annotations on sound clips collected by the TagATune game
- Source codes of the scripts of systems
  - Example (3): TagATune game - the source code of the scripts, and a detailed analysis of the track's structure and musical content

Category 4: Datasets (2)

- Statistics of game players
  - Example (4): the ESP Lite game developed by Chen et al.
- Statistics for task completion
  - Example (5): Ipeirotis et al. gathered all available information from AMT by computing daily statistics for new projects and completed tasks once a day and shared the dataset to the public
- Social tagging dataset
  - Example (6): To extract relationships among tags and resources from the available datasets of two social bookmarking systems, Bibsonomy and GiveALink

Multimedia QoE Assessment

- Quality of Experience (QoE) = User's subjective satisfaction about a service (multimedia content)
- To provide end-user experience, we measure the QoE of multimedia content, e.g, image, voice, video, etc.
- Efficiency vs. Reliability
- Objective evaluation approach
- Subjection evaluation approach

Evaluation Approaches

- Objective Evaluation
  - Cannot capture all the QoE dimensions that may affect users' experiences
  - Cannot include external factors, e.g., quality of headsets, distance between the viewer and the display
- Subjective Evacuation
  - Opinions, e.g., 1=bad, 2=poor, 3=fair, 4=good, and 5=excellent
  - Difficult to define the ordinal scales concisely
  - Difficult to verify users' scoring results

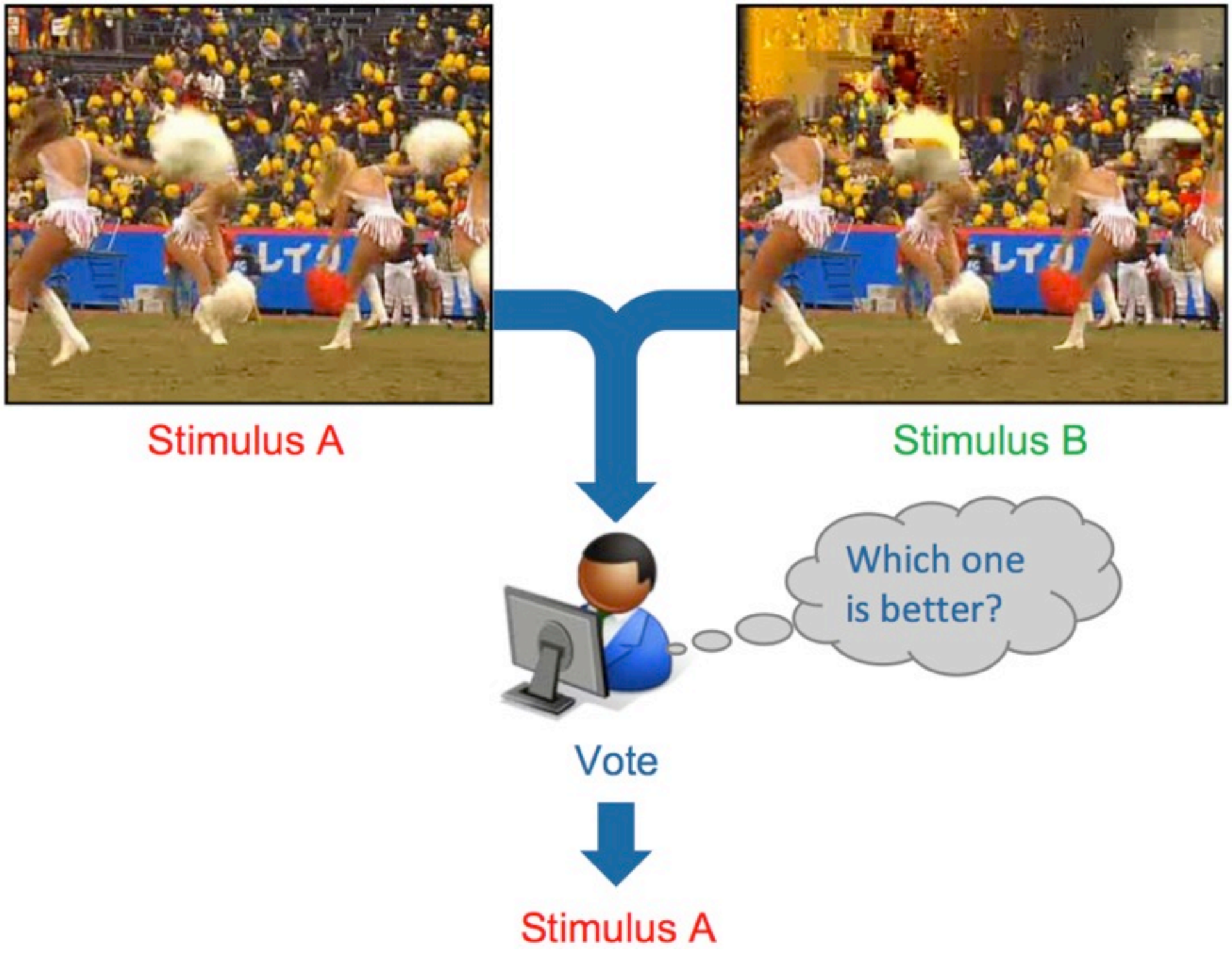
Drawbacks of Subjective Evaluation

- High economic cost
  - Participant payment
- High labor cost
  - Supervision labor
- Physical space/time requirements
  - Transportation cost
  - Laboratory space
  - Difficult to find motivated participants

Crowdsourcing Challenges

- Not every Internet user is trustworthy
- Experiments without supervision so no quality assurance
- Increased variance and bias
- Need to find a way to detect problematic inputs!

Paired Comparison Test



Features of Paired Comparison

- Generalizable across a variety of multimedia applications
- Simple comparative judgement
- Interval scale QoE scores can be calculated
- Verifiable users' feedback

Verification of Users' Inputs

- Transitivity property
  - If A > B and B > C then A should be > C
- Transitivity Satisfaction Rate (TSR)

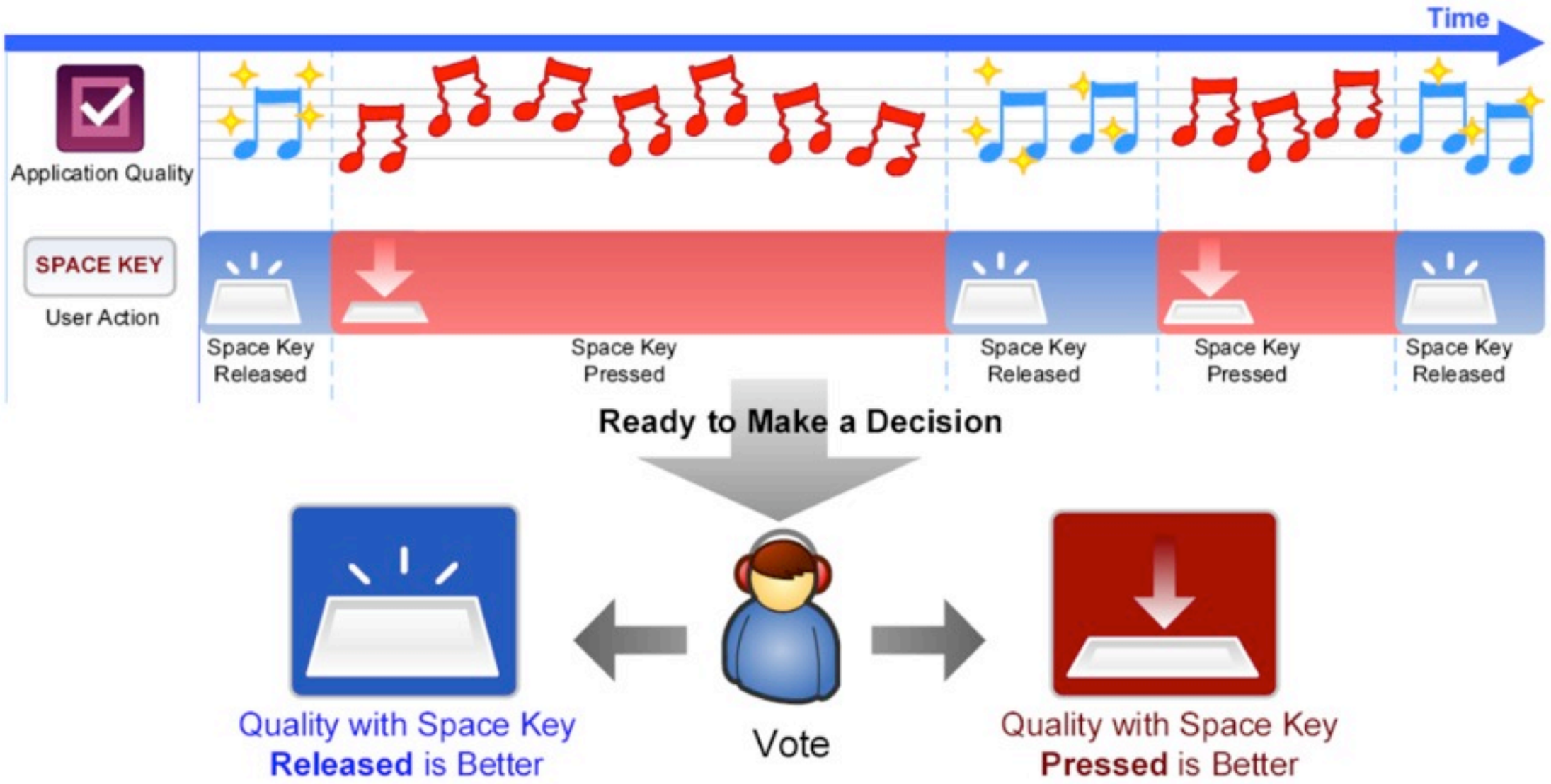
$$\frac{\#\text{ of triples satisfy the transitivity rule}}{\#\text{ of triples the transitivity rule may apply to}}$$

- Detect inconsistent judgements from problematic users
  - TSR = 1 => perfect consistency
  - TSR >= 0.8 => generally consistent
  - TSR < 0.8 => judgement are consistent

Experiment Design

- Suppose our task is to evaluate the effect of n audio processing algorithms, e.g., audio encoding
  - Select an audio clip (source clip) as the evaluation target
  - Apply the n algorithms to the source clip and generate *n* different versions of the clip (test clips)
  - Create an Adobe Flash-based system for users to evaluate the *n* test clips
  - A user need to perform 2 out of *n* paired comparison

Concept Flow of Acoustic QoE Evaluation



Which One is Better?

Participant Source

- Laboratory
  - Recruit part-time workers at an hourly rate of USD $8
- MTurk
  - Post experiments on the Mechanical Turk web site
    - Pay the participant USD $0.15 for each qualified experiment
- Community
  - Seek participants on the website of Internet community with 1.5 million members
  - Pay the participant an amount of virtual currency that was equivalent to USD $0.01 for each qualified experiment

Evaluation of the Framework

- Three participant sources
  - Laboratory
  - Amazon Mechanical Turk
  - Community
- Each with different cost structure
- Compare the cost required by each participant and the data quality produced

- The first **crowdsourcable** QoE evaluation framework
- Users' inputs can be **verified**
  - the transitivity property: A > B and B > C ➔ A > C
  - detect inconsistent judgements from problematic users
- Experiments can thus be outsourced to Internet crowd
  - **lower monetary cost**
  - **wider participant diversity**
  - **maintaining the evaluation results' quality**

Case Study	Experimenter Source	Total Cost (dollar)	# Rounds	# Person	Qualified Rate	Cost / Round (cent)	Time / Round (sec)	Avg. TSR
MP3 Bit Rate	Laboratory	50.97	1440	10	67%	3.54	16	0.96
	MTurk	7.50	750	24	47%	1.00	9	0.96
	Community	1.03	1,470	93	54%	0.07	25	0.96

Chen et al, "A Crowdsourceable QoE Evaluation Framework for Multimedia Content," *Proceedings of ACM Multimedia 2009*.

The http://mmnet.iis.sinica.edu.tw/link/qoe King

Summary

- Crowdsourcing provides a new paradigm and a new platform for scientific research
- New applications, new methodologies, and new businesses are emerging with the aid of crowdsourcing


I think you'll be
delightfully surprised
by the quality of my work
on this assignment.
I crowdsourced it.